The Dissertation Committee for Myriam Lopez Wallace Certifies that this is the approved version of the following dissertation:

Modeling Cross-Classified Data With and Without the Crossed Factors' Random Effects' Interaction

Committee:

S. Natasha Beretvas, Supervisor

Gary Borich

Keenan Pituch

Tiffany Whittaker

Greg Roberts

Modeling Cross-Classified Data With and Without the Crossed Factors' Random Effects' Interaction

by

Myriam Lopez Wallace, B.A.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2015

Dedication

For my husband, Jason, and our son, Rhys Antonio Wallace, born eleven days after my defense.

Modeling Cross-Classified Data With and Without the Crossed Factors' Random Effects' Interaction

Myriam Lopez Wallace, Ph.D.

The University of Texas at Austin, 2015

Supervisor: S. Natasha Beretvas

The present study investigated estimation of the variance of the cross-classified factors' random effects' interaction for cross-classified data structures. Results for two different three-level cross-classified random effects model (CCREM) were compared: *Model 1* included the estimation of this variance component and *Model 2* assumed the value of this variance component was zero and did not estimate it. The second model is the model most commonly assumed by researchers utilizing a CCREM to estimate cross-classified data structures.

These two models were first applied to a real world data set. Parameter estimates for both estimating models were compared. The results for this analysis served as a guide to provide generating parameter values for the Monte Carlo simulation that followed. The Monte Carlo simulation was conducted to compare the two estimating models under several manipulated conditions and assess their impact on parameter recovery. The manipulated conditions included: classroom sample size, the structure of the cross-classification, the intra-unit correlation coefficient (IUCC), and the cross-classified factors' variance component values. Relative parameter and standard error bias were calculated for fixed effect coefficient estimates, random effects' variance components, and the associated standard errors for both.

When *Model 1* was used to estimate the simulated data, no substantial bias was found for any of the parameter estimates or their associated standard errors. Further, no substantial bias was found for conditions with the smallest average within-cell sample size (4 students). When *Model 2* was used to estimate the simulated data, substantial bias occurred for the level-1 and level-2 variance components. Several of the manipulated conditions in the study impacted the magnitude of the bias for these variance estimates. Given that level-1 and level-2 variance components can often be used to inform researchers' decisions about factors of interest, like classroom effects, assessment of possible bias in these estimates is important. The results are discussed, followed by implications and recommendations for applied researchers who are using a CCREM to estimate cross-classified data structures.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Data encountered in social and behavioral sciences can often be hierarchically structured. Shared group environments create a natural clustering structure that introduces dependencies among the observations, which must be appropriately modeled in analyses. A number of educational settings provide examples of hierarchically structured datasets, namely, students nested in classrooms or students nested in schools. Students who come from the same classrooms, or schools, are more likely to share similar experiences and contexts, which creates a dependence among observed outcomes (e.g., test scores). The traditional hierarchical linear model (HLM) is typically used to model nested data structures, with the assumption that they are *purely* hierarchically nested. That is, each lower level unit is associated with only one higher level unit. Purely nested data structures are not always necessarily observed, however, and cross-classification might occur, where the lower level unit is not purely nested in the higher level units. The present study focuses on the cross-classified random effects model (CCREM), an extension of the traditional HLM, which appropriately models non-purely nested, or cross-classified, data structures.

In data structures, a pure hierarchy occurs when lower level units are *clearly* nested in higher level units. For example, students from the same schools can be considered nested within that school. When several schools are located in one neighborhood, these schools can be considered nested within that neighborhood, across several neighborhoods. A pure hierarchical structure occurs when students who attend the same school all live in the same neighborhood.

Purely nested data structures are not always found. To continue with the above example, rarely will every student from a particular neighborhood all attend the same set of schools, nor will every student from one school all live in the same neighborhood. Cross-classification might be found in these scenarios, where students are cross-classified by school and neighborhood. That is, students from the same school may come from different neighborhoods. The traditional HLM cannot be used to model this cross-classification without deleting the students that are not purely nested in schools and neighborhoods or ignoring either the school or neighborhood clustering unit, thereby eliminating the cross-classification (Beretvas, 2008; Goldstein, 2010; Luo & Kwok, 2009; Meyers & Beretvas, 2006). Deleting or ignoring non-purely nested cases from the analyses leads to a loss of power and limits the generalizability of the findings (Beretvas, 2008; Meyers & Beretvas, 2006). The traditional HLM can be extended, however, to properly model the cross-classification that is occurring between the two upper level factors, such as schools and neighborhoods. The present study investigates the cross-classified random effects model (CCREM), an extension of the traditional HLM, which appropriately models non-purely nested data structures, without deleting or ignoring students that are not purely nested.

As with HLM, CCREM models include random effects associated with the lower level, or level-1 unit, and the higher level, or clustering, factors. In CCREM, at least two of the higher level factors will be considered cross-classified, which occurs when the lower level units are not purely nested in these higher level factors. When cross-classified factors occur at the same level, the random effects associated with these factors are also modeled at the same level. For example, when a model includes two cross-classified factors, say, J_1 and J_2, random effects associated with these two factors are included at the same level, with an assumed mean of zero, and variance of τ_{j1} and τ_{j2}, respectively. Additionally, cross-classification at the same level introduces another random effect; the *interaction* of the cross-classified factors' random effects, assumed to follow a normal distribution, with a mean of zero and a variance of $\tau_{j1\ X\ j2}$. Here, this additional random effect represents the interaction between the random effect associated with J_1 and the random effect associated with J_2. If these two crossed factors represent say, classrooms, modeling the interaction of their random effects would allow a researcher to estimate the variability attributed to a student attending a particular combination of classrooms J_1 and J_2. That is, after estimating the random effects' variance associated with J_1 and J_2, a researcher can also estimate the variance associated with the interaction of having attended, or been associated with, classrooms J_1 and J_2. Often is the case, however, that applied researchers utilizing the CCREM will model this variance component $\tau_{j1\ X\ j2}$ as equal to zero, while the variances associated with the crossed factors' random effects (τ_{j1} and τ_{j2}) are estimated (see for example, Attali & Powers, 2009; Lloyd, Li, Hertzman, 2010; Szapocznik et al., 2006). Any variance in the outcome attributed to the random effects' interaction, $\tau_{j1\ X\ j2}$, is not estimated and would then be associated in a different variance component in the model, possibly causing overestimation or underestimation of these components or their standard errors.

The common practice of excluding this variance component in the estimation model (that is, assuming its value to be zero) does require some reasoning. Often, researchers will exclude this component because cross-classified data structures can include small *within-cell* sample sizes, which do not facilitate reliable estimation of this component (Goldstein, 2003; Raudenbush & Bryk, 2002). Small within-cell sample sizes indicates there are not enough level-1 units associated with a particular combination of level-2 cross-classified factors to allow for unbiased estimates of this variance component. For the above school and neighborhood example, this would mean that not enough students are found in a specific neighborhood by school combination that allows for a reliable estimation of $\tau_{j1\ X\ j2}$. Consequently, applied researchers typically set this variance component to zero when modeling their data with the CCREM. To date, however, no simulation paper has been conducted that investigates small within-cell sample sizes and unbiased estimates of the random effects' interaction variance component. Additionally, only one simulation paper has addressed what happens to the other variance components in this type of CCREM model, when this interaction is assumed to be zero (Shi, Leite, & Algina, 2010).

While several methodological papers addressing misspecification in CCREM models can be found (see for example, Meyers and Beretvas, 2006; Luo and Kwok, 2009), at present, only one methodological paper has been found that investigates the cross-classified factors' random effects interaction and its variance component (Shi, Leite, &

Algina, 2010). Shi et al. investigated how different conditions of this variance component affect parameter estimates and associated standard errors. They conducted a simulation study, manipulating several factors in a two-level CCREM, including whether the random effects' interaction variance component was estimated, or assumed to be zero. The study found that its' omission in the estimating model resulted in unacceptable relative parameter bias for the level-2 cross-classified factors' random effects variance components. Shi et al.'s findings indicated that the variance of the random effects' interaction, when assumed to be zero in an estimating model, is redistributed to other variance components located at the same level of the cross-classification in a two-level model.

Shi et al. (2010) only investigated a two-level CCREM. Little is known regarding how the variance of the cross-classified factors' random effects' interaction might be re-distributed when the CCREM model includes a level above the cross-classified level. Previous research has found that in a three-level CCREM, misspecification at the level where the cross-classification occurs, can affect variance components at higher and lower levels (Luo & Kwok, 2009). Specifically, Luo and Kwok found that misspecification at the intermediate, or second, level can affect variance components at higher and lower levels. Misspecification of a three-level CCREM, by exclusion of the cross-classified factors' random effects' interaction at the second level, might result in similar redistribution and overestimation of variance components in levels above and below. To date, no one has specifically addressed the misspecification of this interaction in the context of a three-level CCREM. The present dissertation extends Shi et al.'s (2010) study by including a level above the cross-classification level. A Monte Carlo simulation investigated how excluding estimation of the cross-classified factors' random effects' interaction variance component in a three-level CCREM affected the random effects' variance components and their associated standard errors. Specifically, when the cross-classification occurs at the second, or intermediate, level.

The present study also manipulated the structure of the cross-classification. Luo and Kwok (2009) found evidence to support that the structure of the cross-classification (partial to more complete cross-classification) effects estimation of parameters and their associated standard errors. They did not, however, include or discuss conditions of the cross-classified random effects' interaction. Additionally, in their investigation of the random effects' interaction, Shi et al. (2010) did not manipulate the cross-classification structure, and, therefore, did not compare how estimates were affected across different structures. The current study extended these comparisons to a three-level CCREM, as well as, looking at two real-world types of cross-classification structures.

The present dissertation conducted two studies to investigate the CCREM as discussed above. For the first study, a real world data set with a cross-classification data structure was analyzed, including and excluding estimation of the cross-classified random effects' interaction variance component. This analysis was conducted for demonstration purposes only, and cannot identify which estimates are more similar to the true estimates. The findings in this study were then used for the second study, where a Monte Carlo simulation was conducted to manipulate several conditions that might affect the estimation of this random effects' interaction variance component on parameter recovery. The conditions investigated were: classroom sample size, the structure of the cross-classification, the intra-unit correlation coefficient (IUCC) and the cross-classified factors' variance component values.

Chapter 2: Literature Review

Educational settings are abundant with hierarchically structured data, and the following chapters will employ various examples to demonstrate the utility of multilevel modeling for these settings. Students who attend the same schools, or have the same teacher, share similar experiences and contexts. These clustering structures present dependencies among student level outcomes that must be appropriately modeled. For instance, students who attend the same school might also live in the same neighborhood. Students from the same schools can be considered nested within that school and these schools can be considered nested within neighborhoods, presenting a hierarchically structured data set.

Nesting levels in hierarchical datasets may be comprised of several levels, the present study however will be limited to discussions of two- and three- level units. Level-1 units are units such as individual students who are clustered within second level units, or level-2 units. This second level unit is the group, or cluster, such as a neighborhood, teacher, or classroom, where the level-1 unit is clustered. A dataset with only level-1 and level-2 units can be referred to as a two-level model. Nesting in a data structure can extend beyond level-2. Level-2 units can also be clustered, or nested, within a higher unit, or level-3 unit. Using a simple example, students can be the level-1 unit, and these students can be nested in level-2 units, here, classrooms, which in turn can be nested in level-3 units, here, schools. This example represents a three-level model. For the sake of efficiency, almost all examples presented in this dissertation will refer to three-level models.

HIERARCHICAL LINEAR MODELS

Pure Hierarchical Structures

Hierarchical structures, as those described above, can be clustered in a variety of ways. For example, datasets with students nested in higher level clusters, like classrooms and schools, can be purely hierarchical. Purely hierarchical nesting occurs when lower level units, in this case, students, are *clearly* nested in higher level units, here, classrooms and schools. That is, students in the same classrooms all attend the same school, with each school containing students in a closed set of classrooms. So, for a dataset that includes observations of students, classrooms, and schools, a purely hierarchical dataset would be one with students who only attend one classroom, and these classrooms are only associated with one school. Data structures are not always necessarily purely nested, however, and details related to different nesting data structures will be presented later in the dissertation.

As mentioned above, the traditional hierarchical linear model, HLM, can be used to handle dependencies inherent in purely nested data structures. Using a table format similar to that found in Rasbash and Browne (2001), Table 1 illustrates this example. This dataset would be considered a pure three-level hierarchy if all students (here, *a* through *r*) are nested within several higher level units (here, kindergarten classrooms) at the second level and each set of these units, or kindergarten classrooms, are together associated with only one clustering unit (here, first grade classrooms) at the third level (Raudenbush & Bryk, 2002). For example, kindergarten classrooms 1 and 2 are nested within first grade classroom *I* and only *I*. Each row in Table 1 represents students enrolled in a specific kindergarten classroom, while each column represents students enrolled in a specific first grade classroom.

Table 1

Students Nested in a Pure Three-level Hierarchy

	First grade classroom			
Kindergarten classroom	I	II	III	IV
1	a, b			
2	c, d, e			
3		f, g, h		
4			i, j,	
5			k, l, m	
6				n, o, p, q, r

Note. Letters *a* through *r* represent eighteen students attending 6 kindergarten classrooms and 4 first grade classrooms.

To further clarify using Table 1, a purely nested relationship is found when the row classification (kindergarten classroom) is nested within the column classification (first grade classroom) and all of the individual units (here, students) in a single row fall under a single column (Rasbash & Browne, 2001). For example, students *a* and *b* are clustered in one *cell*, or kindergarten by first grade classroom combination, where they all attended kindergarten class 1 and first grade class I. Students *c*, *d*, and *e* are also clustered in one cell, or kindergarten by first grade classroom combination, in that these students all attended kindergarten classroom 2 and first grade classroom I. There is no deviation from the higher level cluster, meaning that any students who attended one kindergarten classroom *together* also attended first grade classes together, with multiple kindergarten classes per first grade classes.

This same classroom and school example can also be depicted using network graphs (Rasbash & Browne, 2001), as seen in Figure 1, where the lines represent associations of the lower level units (students) with the higher level units (kindergarten and first grade class) and, in the absence of crossed lines, the data can be considered to be purely clustered (Beretvas, 2008).

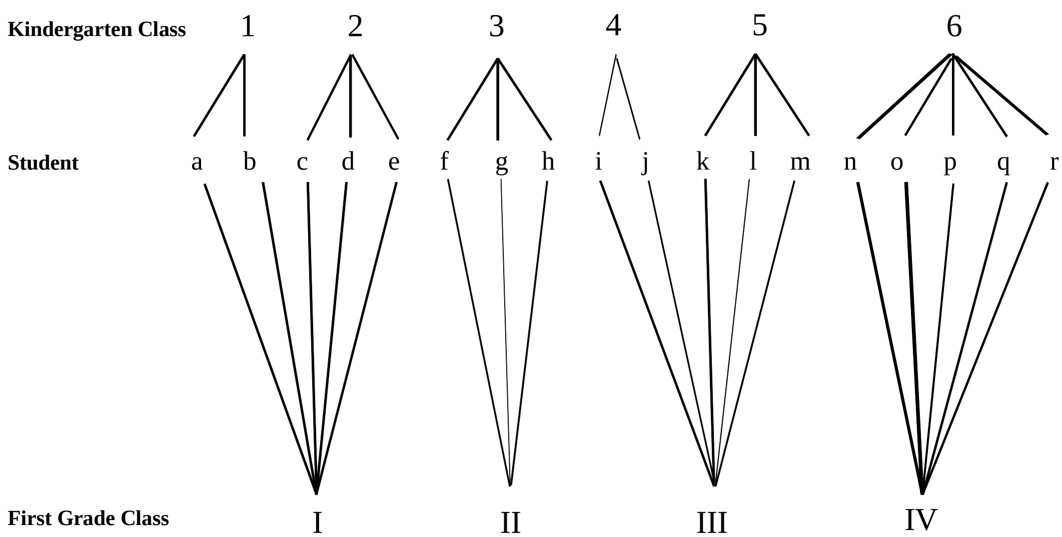

Figure 1. Network Graph of pure three-level hierarchy. Network graph depicting the pure nesting of a three-level hierarchy, where students a through r are purely nested in kindergarten classes, which in turn are purely nested in first grade classes.

HLM Parameterization

The Unconditional Model

The previous school example involving a three-level HLM model, with students nested in level-2 units (here, kindergarten classrooms), and these classrooms nested within level-3 units (here, first grade classrooms), will be used for this review of the HLM formulation. These formulations follow the notation used in Raudenbush and Bryk (2002). Using these notations, the unconditional HLM model, or random intercept model with no predictors at any level, can be represented as follows for level-1:

$$Y_{ijk} = \pi_{0jk} + e_{ijk}, \qquad (1)$$

where Y_{ijk} represents the outcome score of the *i*th student attending kindergarten class *j* and first grade class *k*; π_{0jk} is the mean outcome score for students in kindergarten class *j* and first grade class *k*; and e_{ijk} is the random effect associated with student *ijk,* or the difference between the *i*th student's outcome score and the average outcome score for students in kindergarten class *j* and first grade class *k*. This residual is assumed to have a normal distribution, with a mean of zero and a variance σ^2_{ijk}.

The level-2 equation for the unconditional HLM can be represented as follows,

$$\pi_{0jk} = \beta_{00k} + u_{0jk}, \qquad (2)$$

where β_{00k} represents the mean outcome score for first grade class *k* and u_{0jk} is the random effect associated with kindergarten class *j*. This level-2 random effect is assumed to follow an independent normal distribution, with a mean of zero and variance of τ_π.

For this particular example, the level-3 equation for the unconditional HLM can be represented as follows,

$$\beta_{00k} = \gamma_{000} + r_{00k}, \qquad (3)$$

where γ_{000} is the grand mean outcome score and r_{00k} is the random effect associated with first grade classroom *k*. This level-3 random effect is assumed to follow an independent normal distribution, with a mean of zero and variance of τ_β.

Equations 1, 2, and 3 can be combined and the unconditional model represented by the equation:

$$Y_{ijk} = \gamma_{000} + r_{00k} + u_{0jk} + e_{ijk}\,, \tag{4}$$

where all of the components contributing to the outcome score *Y* can be seen at once. Here, it is evident that the outcome Y_{ijk} is modeled as a function of the grand mean and the random effects associated with the second level unit (here, kindergarten classroom), the third level unit (here, first grade classroom), and the individual.

The fully unconditional model, as seen in Equation 4, is frequently utilized to measure the proportion of unexplained variability in the outcome at each level in the model (Raudenbush & Bryk, 2002). Often, researchers will add predictor variables to the unconditional model to help explain remaining variability. For example, a predictor variable like student gender can be included in a model to help explain some of the variability that exists between students. A model that includes at least one predictor variable is referred to as a conditional model, and is explored further after the following section.

Interclass Correlation Coefficients

The intra-class correlation coefficient (ICC) in HLM provides a measure of outcome variability that is accounted for by the grouping, or clustering, of the data

(Raudenbush & Bryk, 2002). It is calculated as a ratio of between-group, or between-cluster, variance in the outcome over the total variance. That is, the ICC provides the percent of variance in the outcome that occurs between any higher level units. It is a useful measure that captures the degree of clustering, or homogeneity that exists for observations within their cluster, with ICC values closer to one implying that observations in the same clusters are highly similar.

In continuing with the three-level example above, ICCs can be calculated for level-2 and for level-3 using the variance components of all levels. For the example above, the correlation between individuals who attended the same kindergarten classroom, *j*, would be calculated with:

$$ICC_j = \frac{\tau_\pi}{\tau_\pi + \tau_\beta + \sigma^2}, \tag{5}$$

where the numerator contains the variance that is between kindergarten classrooms, τ_π, and the denominator contains the total variance in the dataset: τ_π, τ_β, and σ^2. In addition, the correlation between individuals who are from the same first grade classroom, *k*, can be calculated with:

$$ICC_k = \frac{\tau_\beta}{\tau_\pi + \tau_\beta + \sigma^2}, \tag{6}$$

Where the variance that is between first grade classrooms, τ_β, is in the numerator and the total variance in the dataset, τ_π, τ_β, and σ^2 is in the denominator.

The Conditional Model

Characteristics of the level-1 or level-2 factors may help the researcher explain some of the variability in the outcome. In the previous example (where students are nested within kindergarten and first grade classrooms), a researcher can include a level-1 predictor for students, which would help explain variability between students. In this example, adding a student level predictor, *X*, at level-1 would modify Equation 1 and yield:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} X_{ijk} + e_{ijk}, \tag{7}$$

for the first level. Here, π_{0jk} is the predicted outcome score for kindergarten classroom *j* and first grade classroom *k* when X_{ijk} is zero; X_{ijk} is the student level predictor for student *i* in kindergarten classroom *j* and first grade classroom *k*; π_{1jk} is the change in the outcome score with a one unit change in X_{ijk}; and e_{ijk} is the random effect that is associated with student *ijk*. The random effect, or residual, e_{ijk}, is assumed to be normally distributed with a mean of zero and a variance σ^2_{ijk}.

For level-2 in this example, adding the student level predictor requires the parameterization of π_{1jk} from level-1. Additionally, a predictor variable can also be added at level-2 that may help explain variability between kindergarten classrooms. With the addition of this class-level predictor, *W*, and parameterization of π_{1jk}, the level-2 equation would be:

$$\begin{cases} \pi_{0jk} = \beta_{00k} + \beta_{01k} W_{1jk} + u_{0jk} \\ \pi_{1jk} = \beta_{10k} \end{cases}, \tag{8}$$

for the second level of this example. Here, W_{1jk} is the classroom level predictor for kindergarten classroom *j*; β_{00k} is the predicted outcome score for classroom *k* when X_{ijk} and W_{1jk} are zero; β_{01k} is the change in the outcome score with a one unit change in the classroom level predictor, W_{1jk}, holding all else constant; u_{0jk} is the random effect that is associated with kindergarten classroom *j*, and finally, β_{10k} is the change in the outcome score with a one unit change in X_{ijk}. The random effect, u_{0jk}, is assumed to be normally distributed with a mean of zero and a variance τ_{π}.

For the third level in this example, a predictor can also be added to further explain variability. The level-3 equation here, with an added classroom level predictor, *Z*, would also include parameterization of β_{01k} and β_{10k}, yielding:

$$\begin{cases} \beta_{00k} = \gamma_{000} + \gamma_{001} Z_{1k} + r_{00k} \\ \beta_{01k} = \gamma_{010} \\ \beta_{10k} = \gamma_{100} \end{cases}, \tag{9}$$

where Z_{1k} is the first grade classroom level predictor, γ_{000} is the predicted outcome score when X_{ijk}, W_{1jk} and Z_{1k} are zero; γ_{001} is the change in the outcome score with a one unit change in the predictor, Z_{1k}, holding all else constant; r_{00k} is the random effect that is associated with first grade classroom *k*, γ_{010} is the change in the outcome score with a one unit change in the classroom level predictor, W_{1jk}, holding all else constant, and finally, γ_{100} is the expected change in the outcome score with a one unit change in X_{ijk}, holding all else constant. The random effect, r_{00k}, is assumed to be normally distributed with a mean of zero and a variance τ_{β}.

The equations for the three levels of the conditional model can be combined into one formula:

$$Y_{ijk} = \gamma_{000} + \gamma_{001}Z_k + \gamma_{010}W_{jk} + \gamma_{100}X_{ijk} + r_{00k} + u_{0jk} + e_{ijk}, \qquad (10)$$

where all of the predictors appear in a single equation. Here, all the predictors are modeled as fixed, in that the slopes associated with these predictors do not vary randomly across their respective units. For example, a student level predictor like gender can be modeled as fixed, therefore it is assumed that its coefficient (which captures the relationship between SES and the outcome) does not vary randomly across classrooms.

The conditional model described above does allow for these coefficients to be randomly (or non-randomly) varying at all levels, however, the decision to model the predictors as fixed for this example was done only to simplify presentation. Additionally, the unconditional model can be extended to include several predictors associated with each level (i.e. two predictors at level-1). For more advanced examples of the conditional model that includes random variability among the predictors, as well as additional same-level predictors, see, for example, Raudenbush and Bryk (2002).

Centering

When predictors are added to a model (as done in Equations 7 through 10), they can be utilized in their natural metric or they can be centered in two ways. Centering the level-1 predictor or predictors affects the interpretation of the intercept in the level-1 model

(Raudenbush & Bryk, 2002). For example, in Equation 7 above, π_{0jk} is the predicted outcome score for kindergarten classroom *j* and first grade classroom *k* when X_{ijk} is zero. A value of 0 for the predictor, X_{ijk}, may not always be meaningful, however. When a value of 0 does not make sense for the level-1 predictor, the values of the predictor can be centered in one of two ways: grand mean centering or group mean centering

Grand mean centering requires the value of predictor X_{ijk} to be centered around the grand mean of X_{ijk}. When the value of the predictor is grand mean centered, Equation 7 would be modified to be:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk} - \overline{X}_{...}) + e_{ijk}, \tag{11}$$

where here, π_{0jk} is now the expected outcome score for an individual, *i*, who's value on X_{ijk} is the same as the grand mean, $X_{...}$. According to Enders and Tofighi (2007), grand mean centering is appropriate when the level-2 predictor is of substantive interest, and the level-1 predictor, or predictors, are being controlled for.

Group mean centering requires the value of predictor X_{ijk} to be centered around the mean of its group (group *j* in cluster *k*). When the value of the predictor is group mean centered, Equation 7 would be modified to:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(X_{ijk} - \overline{X}_{.jk}) + e_{ijk}, \tag{12}$$

where, π_{0jk} is now the expected outcome score for an individual, *i*, who's value on X_{ijk} is the same as their group mean, $X_{.jk}$. According to Enders and Tofighi (2007), group mean centering is appropriate when interpretation of the level-1 predictor is of substantive interest to the researcher.

For the present dissertation, interpretation of the predictors is not of interest to the study, and is, therefore, not used nor discussed in detail in the parameterization of the models. Additionally, a researcher may be investigating contextual models. That is, models that explore the relationship between a predictor and the outcome at multiple levels (level-1 and level-2, for example). Researchers interested in contextual models must center their predictors for better interpretation of their outcomes. The current dissertation is not focused on a model with contextual effects and the simulated predictors were not centered. Researchers investigating contextual effects, or who are interested in the effects of predictors at any level, should apply the appropriate centering technique to their data (see, for example, Enders & Tofighi, 2007; Raudenbush & Bryk, 2002).

Model Selection

Using the HLM models described above to fit their data, researchers are able to utilize model information criteria to inform their decisions about the best fitting model, among a set of competing models (Beretvas & Murphy, 2013; Gurka, 2006; Whittaker & Furlow, 2009). That is, researchers utilizing HLM models to fit their hierarchically structured data can obtain model information criteria values that allow them to compare the fit of a group of competing models, such as two conditional models with different predictors, or two models with different random effects. Several criteria are available to compare the fit of HLM models, however, they are not necessarily utilized by researchers, nor are they automatically produced by some of the multilevel software. Here, only the three default information criteria produced by running the PROC MIXED procedure in the latest SAS software (SAS Institute, 2012) are discussed.

One of the more commonly used criteria is the Akaike's (1973) information criteria (AIC), which allows for comparison of nested or non-nested HLM models. The AIC can be calculated with,

$$\mathrm{AIC} = -2LL + 2q, \tag{13}$$

where the -2LL is the deviance statistic for the model and q is the number of parameters estimated. When using AIC to select a better fitting model, a lower AIC value would indicate a better fit.

Some inconsistency issues can arise with the use of the AIC in model fit (Beretvas & Murphy, 2013; Gurka, 2006; Whittaker & Furlow, 2009). A corrected version of the AIC has been proposed, the AICC, which corrects for the inconsistency of the AIC (Hurvich & Tsai, 1989):

$$\mathrm{AICC} = -2LL + 2qN(N - q - 1) \tag{14}$$

where N is the sample size, which can either be the number of level-1 units (N) or can also refer to the number of level-2 units (sometimes referred to as m). Researchers also utilize the Bayesian information criterion (BIC, Schwarz, 1978):

$$\mathrm{BIC} = -2LL + \ln(N)q \tag{15}$$

where *ln* is the natural log and, again, N here can refer to level-1 or level-2 units. The AICC and BIC are often reported in addition to the AIC. As with the AIC, a lower value of the AICC or the BIC would indicate a better fitting model.

While use of model selection with HLM models is still fairly uncommon, it is a useful tool that is available for researchers looking to inform their selection among several competing models (Beretvas & Murphy, 2013; Gurka, 2006; Whittaker & Furlow, 2009). Model information criteria for multilevel models are not limited to purely nested data structures (as presented above) and can be used with more complex data structures that are described in the next section. Thus far, the focus of this chapter has been on purely nested data structures, however, data are not always purely nested. The following section in this chapter will examine more complex structures that may be encountered in educational data.

CROSS-CLASSIFIED STRUCTURES

When working with real datasets, especially in the social sciences, the "purely" nested structures discussed above do not always occur (Rasbash & Goldstein, 1994). Instead, data can often be cross-classified rather than purely clustered multilevel data structures. A cross-classification, as opposed to a pure hierarchy, occurs when there is a lack of clear hierarchy between two higher level clustering factors in multilevel data.

Continuing with the kindergarten and first grade classroom example, rarely, will *every* student from a particular kindergarten class all attend the same first grade class, nor will *every* student from a particular first grade class all come from the same kindergarten class (for a different example of this structure see, for example, Meyers & Beretvas, 2006). Often, kindergarten and first grade classrooms will be "crossed" factors, or units, and students that attend a first grade classroom school might come from several different kindergarten classes, and the kindergarten classroom factor will not be purely hierarchically nested in the first grade classroom factor, nor vice-versa.

This non-hierarchical relationship can be seen in Table 2, where the row classifications (kindergarten class) no longer belong to only one column classification (first grade), as they did in Table 1. Now, these same students are cross-classified by kindergarten classrooms (here, considered Factor 1) and first grade classrooms (here, considered Factor 2). What was previously described as a pure three-level hierarchical structure is now a two-level cross-classified structure, with kindergarten and first grade classrooms now becoming crossed factors at level-2, instead of pure hierarchically nested units at level-2 and level-3.

Table 2

Students Nested in a Cross-classified Data Structure

	First grade classrooms (Factor 2)			
Kindergarten classroom (Factor 1)	I	II	III	IV
1	a, b			
2	c, d		*e*	
3		f, g, h		
4	*i*		j	
5		*l*	k	*m*
6			*n, o, p*	q, r

Note. Italicized letters represent students who do not follow the purely nested data structure.

For example, students *c*, *d*, and *e* all belong to the same Factor 1 unit (kindergarten classroom 2). However, only students *c* and *d* are nested in the same Factor 2 unit (first grade classroom I), while student *e* is now nested in a different Factor 2 unit (first grade classroom III). Factor 1 is no longer purely nested within Factor 2, thereby producing a cross-classified data structure (Meyers & Beretvas, 2006). When cross-classified data are presented in a table format like Table 2, the proportion of empty cells (that is, cross-classification cells that do not contain any level-1 units) can be considered a measure of the degree of cross-classification that exists, such that a greater proportion of empty cells implies a lesser degree of cross-classification (see Luo & Kwok, 2009; Meyers & Beretvas, 2006).

As with the purely clustered relationships, network graphs can also be used to illustrate cross-classified relationships (Rasbash & Browne, 2001). Network graphs of cross-classification, as seen in Figure 2, now illustrate a relationship where there can be "crossing" of the lines connecting some of the higher level clustering units (here, kindergarten and first grade classrooms) with the lower level unit (here, student). This crossing of the lines, which did not exist in the pure hierarchical clusters of Figure 1, illustrates the cross-classified structure found in the cells of Table 2 (Beretvas, 2010). The data structure is no longer a pure hierarchical three-level structure, but rather a two-level data structure where the clustering factors, here, kindergarten and first grade classrooms, are cross-classified.

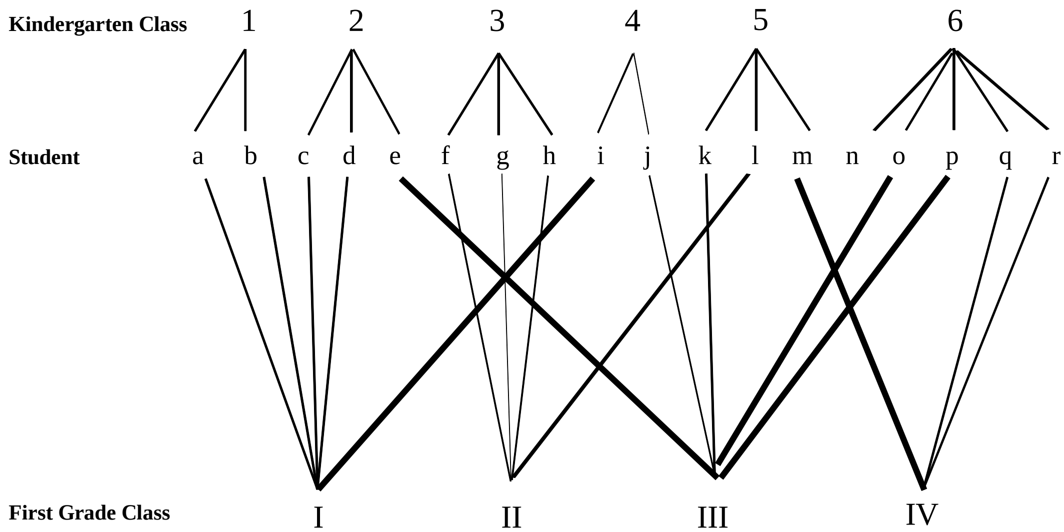

Figure 2. Network Graph of two-level cross-classification. Network graph depicting the cross-classification of students by kindergarten and first grade classrooms.

Modeling Cross-classified Data Structures

Researchers analyzing clustered data that are not purely hierarchical have three choices for handling the data: delete units that lead to the cross-classification, ignore one of the cross-classification factors and model only one of the cross-classified factors, or appropriately model the cross-classification that is occurring using the cross-classified random-effects models (CCREM) (Beretvas, 2008; Goldstein, 2010; Luo & Kwok, 2009, 2012; Meyers & Beretvas, 2006). The first choice, deletion, would involve keeping only level-1 units that maintain the strict hierarchical structure in the higher levels. For the current example, the researcher would delete from analysis any student that did not maintain the strict kindergarten classroom nested in first grade classroom structure. In practice, some applied studies using cross-classified data structures have been found to delete these cases (see, for example, Ainsworth, 2002; De Fraine, Van Landeghem, Van Damme, & Onghena, 2005; McCoach, O'Connell, Reis, & Levitt, 2006). Unfortunately, deleting cases leads to loss of information (decreased power) and limits the generalizability of the results (Meyers & Beretvas, 2006; Beretvas, 2008).

The second option would involve ignoring one of the two cross-classified factors. Applying this method for the current example, the research would choose to either ignore the kindergarten classroom factor or the first grade classroom factor, meaning they would only model students nested in kindergarten classrooms, or students nested in first grade classrooms. Following either of these two procedures allow the researcher to assume purely clustered data and then conduct a traditional HLM analysis.

This second method for handling cross-classified data structures (ignoring one of the crossed factors), is also being utilized in applied research with cross-classified data structures (see, for example, George & Thomas, 2000; Ma & Ma, 2004; Ma & Wilkins, 2002). For the data structures in Table 2 and Figure 2, this approach to cross-classification would require that either the kindergarten classroom classification or the first grade classroom classification be ignored. Ignoring one of these classifications would result in a pure two-level hierarchical structure, with students at level-1 and either kindergarten classroom or first grade classroom at level-2. Not accounting for one of the crossed factors can be considered a misspecification of the model, and can lead to spurious conclusions (Beretvas, 2008) drawn from biased parameter estimates and standard errors of the fixed estimates (Luo & Kwok, 2009; Rasbash & Brown, 2001).

The most appropriate approach to cross-classified data structures would be to properly model these structures using cross-classified random effects models (CCREM). Cross-classified random effects models have been introduced as an extension of the HLM, which can properly handle cross-classified data structures, without having to ignore a classification factor (and misspecify the model) or delete cases (and lose relevant information). Although the CCREM is not as commonly used as the HLM (even with cross-classified data sets), multilevel textbooks often introduce and explain the model (i.e. Goldstein, 1995; Hox, 2002; Raudenbush & Bryk, 2002).

A search in Academic Search Complete, Business Source Complete, EconLit, Health Source, MEDLINE, and PsycINFO, of the terms "CCREM OR Cross-classified OR Crossed factors" and "random effects" resulted in 36 journal articles (not including book chapters and non-peer-reviewed journals) over the last 12 years. These articles describe research conducted across many disciplines including statistics, education, sociology, marketing, agriculture, as well as the medical and psychological sciences. Of these articles, 11 were methodological studies investigating properties of the cross-classified random effects model and 25 were applied studies utilizing the model in their analyses of cross-classified data structures (additional articles were found that were unrelated to the CCREM). Before the relevant methodological literature on the CCREM and its estimation is reviewed, the parameterization of the CCREM will first be reviewed.

CCREM Parameterization

The Unconditional Model

The previous education example with students cross-classified by two level-2 classifications (here, kindergarten and first grade classroom) will be used for this review of the CCREM's formulation. Following Beretvas' (2008), notations for the CCREM formulations in this dissertation use a combination of notations found in Raudenbush and Bryk (2002), Hox (2002), and Rasbash and Browne (2001). In a pure hierarchy, the subscripts *i*, *j*, and *k* can be used to represent each level. For a cross-classified data structure, crossed factors that occur at the same level use the same subscript (here, *j*), contained between parentheses. The subscripts are numbered to differentiate between the two factors, but their ordering is arbitrary.

The unconditional CCREM model (with no predictors at any level) can be represented as follows for level-1:

$$Y_{i(j_1,j_2)} = \beta_{0(j_1,j_2)} + e_{i(j_1,j_2)}, \tag{16}$$

where $Y_{i(j_1 j_2)}$ represents the outcome score of the *i*th student attending kindergarten class j_1 and first grade class j_2; $\beta_{0(j_1,j_2)}$ is the average outcome score for students in kindergarten class j_1 and first grade class j_2; and $e_{i(j_1,j_2)}$ is the random effect associated with student $i(j_1, j_2)$, or the difference between the *i*th student's outcome score and the average outcome score for students in kindergarten classroom j_1 and first grade classroom j_2 and is assumed to have a normal distribution, with a mean of zero and a variance of σ_e^2 (Rasbash & Brown, 2001). The *j* subscripted symbols in parentheses represent the cross-classified factors at the same level (level-2, here).

The level-2 equation for the unconditional CCREM can be represented as follows,

$$\beta_{0(j_1,j_2)} = \gamma_{000} + u_{0j_10} + u_{00j_2} + u_{0(j_1\ X\ j_2)}, \qquad (17)$$

where γ_{000} represents the overall mean outcome score, and u_{0j_10}, u_{00j_2}, and $u_{0(j_1\ X\ j_2)}$ are the random effects associated with the two crossed factors, kindergarten class j_1 and first grade class j_2, and the interaction of the random effects of j_1 and j_2, respectively. The latter random effect, $u_{0(j_1\ X\ j_2)}$, is the variability attributed to the interaction of the crossed factors, j_1 and j_2. All three level-2 random effects are assumed to follow independent, normal distributions, with means of zero and variances of $\tau_{0j_10}, \tau_{00j_2}$, and $\tau_{0(j_1xj_2)}$, respectively.

The equations for both levels can be combined and the unconditional model for CCREM can be represented in one equation:

$$Y_{i(j_1,j_2)} = \gamma_{000} + u_{0j_10} + u_{00j_2} + u_{0(j_1\ X\ j_2)} + e_{i(j_1,j_2)}, \qquad (18)$$

where all of the components contributing to the outcome score Y can be seen at once. Here, it is evident that the outcome, $Y_{i(j_1,j_2)}$ is a function of the grand mean, the random effects associated with the cross-classified factors (here, kindergarten and first grade classrooms), the interaction of these random effects, and the random effect associated with student $i(j_1, j_2)$. Like the HLM model, the unconditional CCREM can be used to observe the amount of variability that remains after accounting for the nesting and crossed-classified structures of the data. Accordingly, researchers can add predictors related to the individual and/or cross-classified factors to further explain variability in the model.

The Conditional Model

As with the traditional HLM, researchers can add level-1 and level-2 predictors to the CCREM unconditional model to help explain variability in the outcome. In the previous example (where students are cross-classified by kindergarten and first grade classrooms), a researcher can include a level-1 predictor for students (to help explain between student variability) and a level-2 predictor for each the two crossed factors to help explain between-kindergarten classrooms and between-first grade classrooms variability. In this example, adding a student level predictor at level-1 and a predictor for each of the level-2 crossed factors would yield the equation:

$$Y_{i(j_1,j_2)} = \beta_{0(j_1,j_2)} + \beta_{1(j_1,j_2)} X_{i(j_1,j_2)} + e_{i(j_1,j_2)}, \tag{19}$$

for level-1 and,

$$\begin{cases} \beta_{0(j_1,j_2)} = \gamma_{000} + \gamma_{010} W_{j_1} + \gamma_{001} Z_{j_2} + u_{0j_10} + u_{00j_2} + u_{0(j_1\ X\ j_2)} \\ \beta_{1(j_1,j_2)} = \gamma_{100} \end{cases} \tag{20}$$

for level-2. Here, $X_{i(j_1,j_2)}$ is the level-1 predictor, whose relationship with *Y*, $\beta_{1(j_1,j_2)}$, is modeled as fixed, or not randomly varying, across the two crossed factors (here, kindergarten and first grade classrooms). W_{j_1} and Z_{j_2} represent the level-2 predictors associated with the cross-classified factors, whose relationships with *Y*, γ_{010} and γ_{001}, respectively, are assumed, in Equation 20, to be fixed across the two cross-classified factors (see Beretvas, 2008 for extensions of this model where the coefficients for crossed-factor predictors are not assumed to be fixed). γ_{000}, is the predicted score on *Y* controlling for *X*, *W*, and *Z*. The random effects, u_{0j_10}, u_{00j_2}, and $u_{0(j_1\ X\ j_2)}$ are still assumed to be normally distributed with a mean of zero, and variances, $\tau_{0j_10}, \tau_{00j_2}$, and $\tau_{0(j_1xj_2)}$, respectively.

Like the unconditional model, the equations for the two levels of the CCREM conditional model can be combined into one equation:

$$Y_{i(j_1,j_2)} = \gamma_{000} + \gamma_{100}X_{i(j_1,j_2)} + \gamma_{010}W_{j_1} + \gamma_{001}Z_{j_2} + u_{0j_10} + u_{00j_2} + u_{0(j_1\ X\ j_2)} + e_{i(j_1,j_2)}, \quad (21)$$

where all predictors are included in the one equation.

The current example was limited to a two-level CCREM. However, cross-classified data structures, like the HLM, can be nested in more than two levels, with cross-classification occurring at any of these levels.

Three-level CCREM

The CCREM, like the HLM, can involve more than two levels. Additionally, crossed factors do not just occur at the second level (as the example above). Cross-classified data structures may be found at the intermediate or top level of a three-level model. Continuing with the example above, cross-classification at the intermediate level may be encountered when students, cross-classified in kindergarten and first grade classrooms, might also include a higher level clustering factor, such as elementary school. Here, the kindergarten and first grade classrooms can be purely nested within the same elementary school. The CCREM used for this type of data structure looks similar to the example above, with additional notation to illustrate the new purely nested factor.

Using the same notation and examples discussed above, and including elementary school as a nesting factor at the third level, the three-level unconditional CCREM model, with the cross-classification occurring at the intermediate level can be represented as follows for level-1:

$$Y_{i(j_1,j_2)k} = \pi_{0(j_1,j_2)k} + e_{i(j_1,j_2)k}, \tag{22}$$

similar to Equation 16 with an additional subscript, *k*, for the level-3 factor, elementary school. The *j* subscripted symbols in parentheses still represent the cross-classified factors at the same level (level-2, here), with an addition of a third-level, *k*, which is not in parentheses, since we are assuming there is pure nesting at this level.

Continuing with the example above, the level-2 equation for the three-level unconditional CCREM is:

$$\pi_{0(j_1,j_2)k} = \beta_{000k} + u_{0j_10k} + u_{00j_2k} + u_{0(j_1\ X\ j_2)k}, \tag{23}$$

similar to Equation 17, where β_{000k} represents the mean outcome score for elementary school *k*, and u_{0j_10k}, u_{00j_2k}, and $u_{0(j_1\ X\ j_2)\mathrm{k}}$ are the random effects associated with the two crossed factors, kindergarten classroom j_1 and first grade classroom j_2 in elementary school *k*, and the interaction of the random effects of j_1 and j_2, respectively. The latter random effect, $u_{0(j_1\ X\ j_2)\mathrm{k}}$, is the variability attributed to the interaction of the crossed factors, j_1 and j_2 within elementary school *k*. All three level-2 random effects are assumed to follow independent normal distributions, with means of zero and variances of $\tau_{0j_10}, \tau_{00j_2}$, and $\tau_{0(j_1xj_2)k}$, respectively.

Lastly, the level-3 equation for the unconditional model is:

$$\beta_{000k} = \gamma_{0000} + r_{000k} \tag{24}$$

where γ_{0000} is the grand mean and r_{000k} is the random effect associated with elementary school *k*, assumed to have a mean of zero and a variance τ_{000k} .The equations for all three levels can be combined and the unconditional model for a three-level CCREM, with cross-classification at the intermediate level is:

$$Y_{i(j_1,j_2)k} = \gamma_{0000} + r_{000k} + u_{0j_1 0k} + u_{00j_2 k} + u_{0(j_1\ X\ j_2)k} + e_{i(j_1,j_2)k}, \qquad (25)$$

where all of the components contributing to the outcome score *Y* can be seen at once. Here, Equation 25 is similar to Equation 18, with the addition of the random effect, associated with the third level factor, *k*.

As with the two-level CCREM, predictors can be added at any level of the three-level unconditional CCREM model to further explain variability in the outcome. Continuing with the previous example, where students are cross-classified by kindergarten and first grade classrooms, and purely nested in elementary schools, a researcher can, for example, include a level-1 predictor for students, level-2 predictors for each the two crossed factors, and/or a level-3 predictor for the elementary school. Extending the example in Equation 21, adding a student-level predictor at level-1 and a predictor associated with each of the level-2 crossed factors, the three-level conditional CCREM would yield the equation:

$$\begin{aligned} Y_{i(j_1,j_2)\mathrm{k}} &= \gamma_{0000} + \gamma_{1000} X_{i(j_1,j_2)\mathrm{k}} + \gamma_{0100} W_{0(j_1,j_2)\mathrm{k}} + \gamma_{0010} Z_{0(j_1,j_2)\mathrm{k}} + \\ & r_{000k} + u_{0j_1 0k} + u_{00j_2 k} + e_{i(j_1,j_2)k} \end{aligned} \qquad (26)$$

where $X_{i(j_1,j_2)k}$ is the level-1 predictor; $W_{0(j_1,j_2)k}$ and $Z_{0(j_1,j_2)k}$ represent the level-2 predictors associated with the cross-classified factors; γ_{0000} is the predicted score on *Y* controlling for *X*, *W*, and *Z;* and the coefficients γ_{1000} γ_{0100} and γ_{0010}, are all assumed to be fixed, not randomly varying. Predictors associated with the level-3 factor can also be added to this equation, and the model can also be extended to include a level-4 and level-5 clustering factor. Additionally, data structures may be observed with cross-classification at the top level. For more detailed modeling of more complicated cross-classified data structures, see, for example, Luo and Kwok (2009).

Intra-unit Correlation Coefficients

For CCREMs the proportion of variability in the outcome that is attributed to factors at each level can be calculated using intra-unit correlation coefficients (IUCCs) (Raudenbush & Bryk, 2002). The IUCC functions similarly to the ICC in Equation 5 and Equation 6 for HLM models. Continuing with the earlier example of students cross-classified by kindergarten and first grade classrooms, and purely nested within elementary schools, and using the variance symbols described above, IUCCs between outcomes of two students (represented as ρ here), can be calculated with:

$$\rho_{j_1 j_2 j_1\ X\ _2 k} = \frac{\tau_{0\,j_1 0k} + \tau_{00\,j_2 k} + \tau_{0\,j_1\ X\ j_2 k}}{\tau_{0\,j_1 0k} + \tau_{00\,j_2 k} + \tau_{0\,j_1\ X\ j_2 k} + \tau_{000k} + \sigma^2_{ij_1, j_2 k}}, \tag{27}$$

for students who attend the same kindergarten and first grade classrooms, which are nested in the same elementary school,

$$\rho_{j_1} = \frac{\tau_{0\,j_1 0k}}{\tau_{0\,j_1 0k} + \tau_{00\,j_2 k} + \tau_{0\,j_1\ X\ j_2 k} + \tau_{000k} + \sigma^2_{ij_1, j_2 k}}, \tag{28}$$

for students who attend the same kindergarten class, j_1, but attended different first grade classrooms, where both are nested in the same elementary school,

$$\rho_{j_2} = \frac{\tau_{00j_2k}}{\tau_{0j_10k} + \tau_{00j_2k} + \tau_{0j_1\ X\ j_2k} + \tau_{000k} + \sigma^2_{ij_1,j_2k}} \qquad (29)$$

for students who went to the same first grade classroom, j_2, but attended different kindergarten classrooms, where both are nested in the same elementary school.

CCREM Random Effects' Interaction

The random effect interaction between the cross-classified factors, $u_{0(j_1\ X\ j_2)}$ for the two-level CCREM and $u_{0(j_1\ X\ j_2)\mathrm{k}}$ for the three-level CCREM above (see Equation 21 and Equation 26, respectively), is not typically included in most applied CCREM analyses (see for example, Attali & Powers, 2009; Lloyd, Li, Hertzman, 2010; Szapocznik et al., 2006). The variance for this random effect, $\tau_{0(j_1 x j_2)k}$, is typically assumed to be zero and, therefore, not included in the estimating model. This random effect, $u_{0(j_1\ X\ j_2)\mathrm{k}}$, is described as the interaction between the two cross-classified factors' random effects. Continuing with the example above, the variance component, $\tau_{0(j_1 x j_2)k}$, is the variance associated with the interaction of first grade classroom and kindergarten classroom. A value of nonzero for this variance would imply that the effect of attending a first grade classroom on a student's outcome is not the same for students coming from kindergarten classrooms with different effects. So, for example, if the outcome of interest was reading comprehension, the effect of a high performing first grade classroom on reading comprehension is not the same for students coming from different kindergarten classrooms. To align with previous literature, the random effect interaction will be discussed in the context of a three-level model, as seen in Equation 26. However, the term and its parameterization are assumed to be almost identical for both a two-level and a three-level model.

Often, the variance of this random effect, $\tau_{0(j_1 x j_2)k}$ is assumed to equal zero (thereby "dropping" $u_{0(j_1\ X\ j_2)k}$ from the estimating model) while the remaining random effects' variances are estimated. Chapters discussing CCREM (see, for example, Beretvas, 2008; Beretvas, 2010; Hox, 2010; Raudenbush & Bryk, 2002) include the interaction term, $u_{0(j_1\ X\ j_2)k}$, in their presentation of the CCREM, however, the variance of the interaction term is assumed to be zero and not estimated because small within-cell sample sizes do not facilitate valid estimation of the variance component, $\tau_{0(j_1 x j_2)k}$ (Goldstein, 2003; Raudenbush & Bryk, 2002). To clarify, small within-cell sizes (or crossed factor by crossed factor combinations) are not uncommon in applied research and, therefore, the true value of the variance component, $\tau_{0(j_1 x j_2)k}$, may not be well recovered. For the example used above, that would mean there are not enough students attending a specific kindergarten classroom, j_1, and first grade classroom, j_2, combination that allow for reasonable estimation of this variance parameter's value. Specifically, small within-cell sample sizes do not allow $\tau_{0(j_1 x j_2)k}$ to be easily estimated separately from the estimation of $\sigma^2_{\mathrm{i}(j_1 x j_2)k}$, the variance component associated with the level-1 residual, $e_{i(j_1 x j_2)}$ (Raudenbush & Bryk, 2002; Beretvas, 2008). Consequently, the variance $\tau_{0(j_1 x j_2)k}$ is typically set to zero in applied research involving use of CCREM (including in the calculation of the IUCCs).

Statistical software, such as SAS (SAS Institute, 2012), can be used to estimate the

CCREM using its PROC MIXED procedure, however, the default for SAS PROC MIXED is to set the variance of the random effects interaction term to zero. SAS PROC MIXED can be programmed to *allow* for the estimation of $\tau_{0(j_1 x j_2)k}$ although none of the CCREM applied studies gathered from the earlier literature search estimated this term. Three applied articles did mention and define the random effects' interaction, $u_{0(j_1\ X\ j_2)k}$, in the context of their study, but assumed that the associated variance, $\tau_{0(j_1 x j_2)k}$ was zero, recognizing the presence of small within-cell sizes in their respective data and identifying the estimation problems this is believed to cause (Attali & Powers, 2009; Lloyd, Li, Hertzman, 2010; Szapocznik et al., 2006).

Methodological studies regarding the properties of the random effects' interaction term, $u_{0(j_1\ X\ j_2)k}$, are sparse, supporting the need for further investigation to explore how assuming the variance of this term, $\tau_{0(j_1 x j_2)k}$, is zero might affect estimates of the remaining parameters and associated standard errors in the model. Raudenbush and Bryk (2002) explain that when within-cell sizes are too small, the variance term, $\tau_{0(j_1 x j_2)k}$, is confounded with the level-1 variance component, $\sigma^2_{i(j_1 x j_2)}$, making it difficult to estimate. Simulation studies have not explicitly investigated how small within-cell sample size effect estimation of this term. In their chapter on cross-classified data structures, Raudenbush and Bryk did note that their example dataset had "very small" within-cell sample sizes, with the number of units in each cell being as low as one. No methodological comparisons were conducted in their description, however, and no guidance was provided as to when within-cell sample sizes might be large enough for reasonable parameter recovery.

While a number of methodological papers have explored model misspecification of CCREM, only one methodological paper has been published that has specifically investigated the term, $u_{0(j_1 \ X \ j_2)k}$, in the context of a CCREM, and the effect that assuming its variance, $\tau_{0(j_1 x j_2)k}$, is zero may have on estimation of parameters and standard errors (Shi, Leite, Algina, 2010). Before describing the Shi et al. paper, previous methodological research focused on model misspecification for CCREM will be reviewed.

CCREM Methodological Studies

Fielding (2002) investigated the effects of ignoring cross-classification using a three-level model for a large data set. This real data set contained a student level achievement outcome at level-1 nested within two cross-classified factors at level-2 (teaching group and student), nested within institution at level-3. In the main analysis, Fielding modeled the data in two ways: 1) inappropriately, as a pure three-level hierarchy, ignoring the "student" factor and 2) appropriately, as a three-level CCREM, recognizing the cross-classification at level-2. Fixed effect estimates were virtually identical under both models. However, level-2 variance estimates differed. When one of the cross-classified factors was ignored at level-2 (student), estimates of the level-1 variance increased, while estimates of the variance for the non-ignored cross-classified factor (teaching group) decreased slightly. The variance at level-3 was not affected. For the real dataset analysis, inappropriate modeling of cross-classified data structures, with cross-classification at level-2, resulted in apparent overestimation of the level-1 variance and underestimation of the level-2 variance. These findings make sense, because if the ignored factor's variance is not zero and, if it is assumed to be zero, then this variance must be redistributed to another variance component in the model. With regards to the interaction term, $u_{0(j_1\ X\ j_2)k}$, Fielding did not include nor mention this term. Because Fielding's methodological investigation did not include a simulation study, the findings warrant further research, as it is unclear which parameter estimates might be closer to the corresponding true values.

Meyers and Beretvas (2006) compared the performance of three methods for analyzing real and simulated cross-classified data structures with two crossed factors, including the: a) CCREM, b) use of HLM after deleting some cross-classified cases (only for the real data structure example, not for the simulation investigation), and c) use of the HLM after ignoring one of the two level-2 cross-classified factors. Meyers and Beretvas looked at a simple two-level model, where level-1 data was nested in two level-2 cross-classified factors (F1 and F2). In the simulation study that the authors conducted, they looked at the parameter estimates, associated standard errors and fit for two of the three methods described above, under conditions of: differing correlations between the residuals of F1 and F2, degree of cross-classification, varying sample sizes, and magnitude of the intra-unit correlation coefficients (IUCCs). Overall, they found that while parameter estimates for the fixed effects were not different across the methods [corresponding with Fielding, (2002)], the associated standard errors for these effects were underestimated when the cross-classified data were treated as purely hierarchical. Not using the CCREM also resulted in *overestimation* of the level-1 residual variance and of the variance of the residuals for the cross-classification factor that was not ignored, with the latter finding contradicting some of Fielding's results. Standard errors were found to be more negatively biased when the true correlation was zero. Additionally, they compared two information criteria, the AIC and BIC, for both the CCREM and the incorrect HLM model and found that the BIC correctly identified the CCREM as the better fit 100% of the time. The AIC was more likely to correctly select the CCREM when the correlation between the residuals was zero and when the IUCC was larger.

Like Fielding, Meyers and Beretvas did not model the random effects' interaction term, $u_{0(j_1 \ X \ j_2)k}$, however, they did acknowledge the term and mentioned that the variance of this term is not usually estimated well (2006). Consequently, the variance of this term was assumed to be zero for both generating and estimating models in their study, and the effect of this assumption were not assessed and its estimation was not evaluated.

Luo and Kwok (2009) extended the study by Meyers and Beretvas (2006) to investigate why some research shows contradictory findings where ignoring one cross-classification factor may lead to *underestimation* of the remaining cross-classification factors' random effects' variance (as was found in Fielding, 2002). The authors' extension included a three-level model, instead of a two-level model, to explore whether the *level* at which the cross-classification occurs may affect bias. For example, if the cross-classification occurs at the highest level in a three-level model then ignoring the cross-classification may affect bias differently than if the cross-classification occurs at the intermediate, or second, level.

Luo and Kwok also extended Meyers and Beretvas' (2006) study by looking at the distribution of the cross-classification. Where Meyers and Beretvas looked at the *number* of empty cells as a marker of cross-classification for one of their conditions, Luo and Kwok (2009) revised this condition by investigating three *structures* of the cross-classification - complete, partial, and hierarchical cross-classification. That is, Luo and Kwok manipulated the cross-classification structure of the data, by changing the probability that a level-1 unit has of belonging to any cluster of the cross-classified factors in higher levels. In the complete cross-classification condition, for example, all units in a cluster of one crossed factor have a roughly equal probability of being affiliated with *any* cluster of the other crossed factor. So if students are cross-classified by school and neighborhood, a complete cross-classification would indicate that students in a particular neighborhood have an equal probability of attending any school in the dataset, and students from any school can live in any neighborhood.

Partial cross-classification, on the other hand, is viewed as more realistic [according to Luo and Kwok (2009)], and is found when units in one crossed-factor cluster are only associated with *some* of the clusters in the other crossed-factor. For example, if students are cross-classified by school and neighborhood, students living in a specific neighborhood will likely attend certain schools, increasing their probabilities of association with certain schools, and the probabilities are no longer equal, as they were in the complete cross-classification condition. While Luo and Kwok did extend these specific conditions of

Meyers and Beretvas' (2006) study, the authors did not manipulate true IUCC values nor the correlation between the two cross-classified factors' random effects. Luo and Kwok also did not examine the effects of using the HLM-deletion method for handling three-level cross-classified data.

With a simulated cross-classified dataset, Luo and Kwok (2009) compared parameter estimates and standard errors for fixed and random effects variance components for: 1) a correct model (CCREM) with cross-classified factors at the appropriate level, 2) a misspecified model (HLM), where one cross-classified factor (F1) was almost fully nested with the other cross-classified factor (F2), and this other factor, F2, was not modeled and 3) a misspecified model (HLM), where the other cross-classified factor, F2, was almost fully nested within F1, and F1 was not modeled. The second and third conditions differed in which cross-classified factor, either F1 or F2 was almost fully nested in the other and which factor was ignored. This was done for two data structures nested in three-levels, one with the cross-classified factors found at level-3 (the top clusters) and one with the cross-classified factors found at level-2 (the intermediate clusters).

For misspecified HLM models with intermediate level cross-classification, slight positive bias was found for the level-1 residuals' variance, which increased as the data structure became more cross classified (Luo & Kwok, 2009). Conversely, positive bias was found for the variance component of the remaining, non-ignored cross-classified factor, supporting the findings in Meyers and Beretvas (2006), and this bias decreased as the dataset became more cross-classified. Positive bias for the level-3 variance component was only found for the misspecified model where F1 was almost fully nested within F2. Overall, Luo and Kwok found that ignoring one of the factors in a cross-classification results in its variance being distributed to random effects' variances at other levels, resulting in overestimation of the remaining variance components at all three levels. The overestimation also seemed to be affected by the structure of the cross-classification. Unlike Fielding (2002), however, Luo and Kwok did find that estimation of level-3 variance components are affected by misspecification of a cross-classified dataset, when the cross-classification occurs at the third level.

In addition to several simulation studies, Luo and Kwok (2009) presented formulas for the variance components of both the CCREM and a misspecified HLM model, which illustrated what should happen to the variance of the ignored cross-classified factor in a misspecified HLM model. Based on these formulas, in a three-level model, with cross-classification at the intermediate level, ignoring a cross-classified factor at the intermediate level should produce overestimated level-1 and level-3 variance components, and underestimated variance components of the remaining cross-classified factors. However, the simulation results indicated that the remaining cross-classified factor's variance was *over*estimated. The authors did not specifically address these differing results and further research should investigate the properties of this variance component under different conditions. Importantly, the structure of the cross-classification was found to affect where and how the variance of the ignored factor's random effects is redistributed. Luo and Kwok also found that standard errors associated with the fixed effect coefficient estimates were found to be underestimated for the predictor variables associated with the factor that was ignored.

Like Meyers and Beretvas (2006), Luo and Kwok (2009) did not include the random effect interaction term, $u_{0(j_1 \ X \ j_2)\mathrm{k}}$, in their generating nor estimating models. However, no explanation or mention of this term was provided. Overall, Luo and Kwok's results supported the findings in Meyers and Beretvas that ignoring one of the cross-classification factors is not a viable option if the researcher wants to reduce bias in the estimation of variance components and standard errors. However, because the random effects' interaction term, $u_{0(j_1 \ X \ j_2)k}$, was not included in the models explored in these studies, conclusions could not be drawn about the effects of its omission on parameter and standard error recovery.

Recently, Luo and Kwok (2012) conducted another study comparing the performance of HLM versus CCREM when handling cross-classified data structures in longitudinal datasets. Longitudinal studies involve repeated measures of an outcome and can often entail cross-classification issues caused by student mobility. For example, if repeated measures are collected for students who are nested in school clusters, the data structure may not necessarily be a three-level pure hierarchy, with repeated measures purely nested within students and students nested in schools. Student mobility may create a cross-classified data structure, where repeated measures are nested within students, but each student might not necessarily be associated with only one specific school. Luo and Kwok investigated the impact of model misspecification for this type of growth data.

Luo and Kwok compared the performance of HLM and CCREM for real and two simulated cross-classified data structures. They looked at coverage (of the confidence interval estimates) and relative bias for estimates of fixed effects, variance components and standard errors, while manipulating the following conditions: a) the total number of schools in the dataset, b) the number of students per school at the first measurement occasion, c) the mobility rate, where students could switch schools between the first and second measurement occasions, d) the variance and covariance of student random effects, associated with the student's intercept and slope parameters (medium and small size), and e) the variance of school random effects (small and medium). A second simulated study manipulated the student mobility condition to mimic more realistic situations, where students switch schools randomly and multiple times. This was similar to Luo and Kwok's (2009) cross-classification conditions, where cross-classification was either complete, partial, or hierarchical.

Overall, results were similar to those of Meyers and Beretvas (2006) and Luo and Kwok (2009) in that misspecification of the model (where cross-classified data was treated as purely hierarchical) resulted in biases in the variance components and standard error estimates of the random effects. As expected, the fixed effects estimates were relatively unaffected, although the associated standard errors showed slight bias. Specifically, under different patterns of student mobility, misspecification by treating a cross-classified data structure as purely hierarchical resulted in overestimation of the variance component at the same level where the cross-classification occurred. Similar to findings by Luo and Kwok

(2009), misspecifying a cross-classified data structure, with a mobility pattern that is similar to *partial* cross-classification, produced overestimation of the variance components associated with the ignored factor (which occurred at level-2) and the covariance of the student random effects associated with the intercept and slope at level-2. They also found that the level-1 residual variance was unaffected.

The findings here support Meyers and Beretvas (2006) and Luo and Kwok (2009), where ignoring a cross-classified factor (thereby misspecifying the model) results in inaccurate estimation of variance components, standard errors, and thus spurious conclusions about the variability that exists in the data structure. Once again, however, the random effects' interaction term, $u_{0(j_1\ X\ j_2)k}$, was not included nor discussed, and its estimation and effects were not explored.

Shi, Leite, and Algina (2010) conducted the only methodological study that investigated the crossed factors' random effects interaction term, $u_{0(j_1\ X\ j_2)}$, and how assuming its variance, $\tau_{0(j_1 x j_2)}$, is zero might affect parameter and standard error estimates. The study used real and simulated cross-classified data structures to investigate the possible bias that can occur when $\tau_{0(j_1 x j_2)k}$ is assumed to be zero and not included in the estimating model. They used two generating CCREM models, both with students cross-classified by two factors (such as middle school and high school) at level-2. Both generating models were two-level models, and included a level-1 predictor and two level-2 predictors (one for each cross-classification factor). Only the level-1 intercept was modeled as varying across the level-two factors. The slope for the level-1 predictor was assumed fixed, as were the slopes for the level-2 predictors. This yielded equations similar to those seen in Equation 19 for level-1 and Equation 29 for level-2 [except that the variance of $u_{0(j_1 \; X \; j_2)}$ was set to zero in one of the generating models].

The only difference between the two generating models was that under one model, the variance of the random effects' interaction, $\tau_{0(j_1 x j_2)}$, was generated to be *non*-zero and under the other model, the same variance was set to zero. Each random effect [$u_{0 j_1 0}$, $u_{00 j_2}$, and $u_{0(j_1 \; X \; j_2)}$] was sampled independently from a normal distribution, with a mean of zero, and respective true variance [$\tau_{0 j_1 0}, \tau_{00 j_2}$, and $\tau_{0 j_1 x j_2}$]. Covariances between the cross-classified factors' residuals were assumed to be zero (Beretvas, 2008; Goldstein, 2003), although real data may not match this assumption (Shi et al., 2010). In summary, data were generated from two identical models, except that one of the models included the variance term, $\tau_{0(j_1 x j_2)}$, while the other model set it to zero.

Shi et al. (2010) investigated estimates of parameters and standard errors under conditions manipulating:

a) the correlation between the residuals for the two cross-classified factors (zero and non-zero correlation), which replicates what is usually assumed about these correlations – that they are zero – and what is likely found with real data (i.e., that they are non-zero) according to Shi et al.,

b) the number of middle schools "feeding" into each high-school, (each school that sends a student to another school is referred to as the *feeder* school and the corresponding school that receives the student is the *receiver*),

c) the true value of the middle school and high school intra-unit correlation coefficients (small and moderate), and

d) the variance, $\tau_{0(j_1xj_2)}$, which was either zero or nonzero in the generating model.

Overall, Shi et al. (2010) found that when the *generating* model included a nonzero $\tau_{0(j_1xj_2)}$ value and a zero correlation between the cross-classified factors' residuals, and the *estimating* model assumed $\tau_{0(j_1xj_2)}$ was zero, then the estimates of the level-2 variance components, τ_{0j_10} and τ_{00j_2}, were positively biased. With the same generating model conditions, when the *estimating* model assumed $\tau_{0(j_1xj_2)}$ was zero, estimations of the level-1 variance component, σ_e^2, were not biased. This would indicate that when no correlation exists between the cross-classified factors' residuals and when the value of $\tau_{0(j_1xj_2)}$ is non-zero, then the assumption that $\tau_{0(j_1xj_2)}$ is zero in the estimating model leads to bias in the variance components of level-2, where the cross-classification occurs. This supports some findings from Luo and Kwok (2009), that misspecification of a CCREM by not estimating a cross-classified factors' variance component can lead to overestimation of the variance components at the same level of the cross-classification. Shi. et al.'s study only included two levels, however, which does not allow for investigation of how this same type of misspecification (assuming $\tau_{0(j_1xj_2)}$ is zero when it is not zero) might affect variance components at other levels, specifically, higher levels, where bias has been found to occur (Luo and Kwok, 2009).

Shi, Leite, and Algina (2010) also found that when the generating model included a value of non-zero for $\tau_{0(j_1 x j_2)}$ *and* a non-zero correlation between the residuals for the two cross-classified factors, but the estimating model assumed that both were zero, then the worst positive bias was found in the estimates of the level-2 cross-classified factors' variance components, τ_{0j_10} and τ_{00j_2}. This provides some evidence that the potential correlation in the crossed factors' residuals should also not be ignored when estimating a CCREM. While most current, canned multilevel modeling software does not allow for estimation of this correlation, including it in the generating model does allow the researcher to observe how values of this correlation might bias parameter and standard error estimates. Last, the authors found that fixed effects estimates and their standard errors were not affected by model misspecification (assuming $\tau_{0(j_1 x j_2)}$ is zero when it is not zero).

STATEMENT OF PURPOSE

While the methodological studies summarized here have contributed essential findings, only Shi et al.'s (2010) study has specifically explored CCREM models that include estimations of the random effects interaction variance component, $\tau_{0(j_1 x j_2)}$. The current study is designed to extend Shi et al.'s study, guided by the findings in Luo and Kwok (2009) concerning intermediate level cross-classification in three-level models. Shi et al. found that the assumption that $\tau_{0(j_1 x j_2)}$ was zero in the estimating model (when it was truly non-zero) resulted in overestimation of the level-2 cross-classified factor variances, τ_{0j_10} and τ_{00j_2}. This supports previous findings that misspecification of cross-classified factor models (especially when one of the crossed-factors is ignored) overestimates variance components at the same level that the cross-classification occurs (Meyers & Beretvas, 2006; Luo & Kwok, 2009). It is reasonable to believe, then, that misspecification by assuming that the random effects' interaction variance, $\tau_{0(j_1xj_2)}$, is zero, when it is truly not, would also affect bias of variance estimations in the model. Specifically, this type of misspecification might affect variance estimations at the same level where the cross-classification is occurring, and, further, affect variance estimations at other levels of the model, as was found for Luo and Kwok (2009).

The current study investigated the effects of assuming $\tau_{0(j_1xj_2)}$ is zero, when it is not zero (as investigated in Shi. et al.) and extended the two-level design by adding a third level to a cross-classified data structure, with cross-classification at the intermediate level and pure nesting at level-3, as was done in Luo and Kwok, (2009). Because Luo and Kwok found that misspecification of a CCREM model with cross-classification at the intermediate level may overestimate both the level-2 and level-3 variance components, it is reasonable to believe that misspecification of a CCREM by setting the variance of $u_{0(j_1\ X\ j_2)}$ to zero would result in a similar pattern of overestimation. By including a third level in the model (that does not contain a cross-classification at that level), the present study investigated how setting the variance of $u_{0(j_1\ X\ j_2)}$ to zero affected estimation of random effects' variance components and their associated standard errors.

Shi et al. found that the variance of $u_{0(j_1 \, X \, j_2)}$ was "redistributed" into other variance components, when it was incorrectly missing from the estimating model. Specifically, the variance components at the same level of the cross-classification, τ_{0j_10} and τ_{00j_2} were found to be *overestimated* when $\tau_{0(j_1xj_2)}$ was nonzero, but assumed to be zero. This warranted further investigation into what characteristics of the cross-classified factors and their respective variance components might influence the "redistribution" of $\tau_{0(j_1xj_2)}$ when it is generated, but not estimated. Because Luo and Kwok (2009) found that ignoring an existing crossed factor resulted in the redistribution of its' variance into both higher and lower levels, a similar pattern might be observed when the variance, $\tau_{0(j_1xj_2)}$, was generated to be nonzero, but not included in the estimating model. The values of the variance components, τ_{0j_10} and τ_{00j_2} may influence how the variance for their interaction, $\tau_{0(j_1xj_2)}$, is redistributed when it is set to zero in an estimating model. Meyers and Beretvas (2006) looked at IUCC values of .05 and .15, which were calculated by fixing the values of τ_{0j_10} and τ_{00j_2} to .0556 in one condition and .2143 in another condition. Shi et al. (2010) replicated these IUCC values, but included fixing the value of the variance component, $\tau_{0(j_1xj_2)}$ to also be .0556 and .2143. The current study extended these conditions by looking at different IUCC values that followed findings from a real world data analysis that was conducted. Additionally, the current study manipulated the relationship between the two level-2 cross-classified variance components, τ_{0j_10} and τ_{00j_2}, and the variance of their interaction, $\tau_{0(j_1xj_2)}$. The present study did not assume that these three values would be equal to each other, as was assumed for Shi et al. The study investigated the extent to which an unequal (and therefore, smaller) variance value for $\tau_{0(j_1xj_2)}$ might influence parameter and standard error estimation. That is, how was the $\tau_{0(j_1xj_2)}$ variance redistributed if it was not included in the estimation of the model, and was unequal to τ_{0j_10} and τ_{00j_2}.

Additionally, in their investigation of the crossed factors' residuals' interaction effect, Shi et al. (2010) did not investigate different within-cell sample sizes. Their sample sizes were a result of the number of middle schools feeding into a high school (where a two-feeder condition would result in larger within-cell sample sizes than a three-feeder condition). Applied research and multilevel textbooks have stated that small within-cell sample sizes make estimation of $\tau_{0(j_1xj_2)}$ problematic. Using several values of classroom sample size and two different feeder patterns, the current study manipulated the average number of level-1 units within each cell to investigate several within-cell average sample sizes and their effects on unbiased estimations of $\tau_{0(j_1xj_2)}$.

Luo and Kwok (2009) found evidence to support that the distribution of cross-classification (that is, partial to more complete cross-classification) affected estimation of parameters when a cross-classified factor was not correctly estimated. It is reasonable to expect similar results to Luo and Kwok's (2009) findings for misspecification of cross-classified data structures, where bias differed as data structures became more cross-classified. The current study extended these comparisons to a three-level model and compare estimates for different distributions of cross-classification.

For the present dissertation, two studies were conducted to investigate the estimation of the crossed factors' random effects' interaction, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ for three-level CCREM models. For both of these studies, results for two different estimating models were compared: one model included the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, and the second model assumed the value of this variance component was zero. For the first study, parameter estimates and their associated standard errors were compared for both estimating models, when applied to a real world data set, the STAR data (STAR Project; Achilles, C.M. et al., 2008). This analysis allowed for the demonstration of where potential differences might occur, but did not identify which estimates are closer to truth. Additionally, some of the results for this analysis served as a guide to provide generating parameter values for a second study. For the second study, a Monte Carlo simulation was conducted that compared the same two estimating models described above, but allowed for investigation of several manipulated conditions and their impact on parameter recovery. The conditions included: classroom sample size, the structure of the cross-classification, the intra-unit correlation coefficient (IUCC), and the cross-classified factors' variance component values (equal and unequal).

Chapter 3: Method

The first study was conducted using a real data set from the Student/Teacher Achievement Ratio (STAR) Project funded by the Tennessee General Assembly (Achilles, C.M. et al., 2008). This study was used to investigate differences in parameter estimates and associated standard errors between two three-level CCREM models: one that included estimation of the random effects' interaction variance, $\tau_{0j_1 \, x \, j_2k}$, and a model that assumed the value of $\tau_{0j_1 \, x \, j_2k}$ was zero and did not estimate it. Using some the estimates from this first study as a guide, a Monte Carlo simulation was then conducted, which allowed investigation of parameter recovery for the random effects' interaction variance, $\tau_{0j_1 \, x \, j_2k}$, and the other variances, under several conditions, including model misspecification.

STUDY 1 - REAL DATA ANALYSIS

The first study provides an example of real world data with a three-level nesting structure, including cross-classification at level-2, similar to the data investigated in Luo and Kwok (2009). The study compared parameter estimates and associated standard errors for two estimating models: a CCREM model that included the estimation of the random effects' interaction variance, $\tau_{0j_1 \, x \, j_2k}$, and a misspecified model that assumed the value of $\tau_{0j_1 \, x \, j_2k}$ was zero. The purpose of this investigation was two-fold: 1) to illustrate an example of estimation differences for two versions of the CCREM to assess potential differences and 2) to provide generating model parameter values for the Monte Carlo simulation in the second analysis that would resemble means, standard deviations, and parameter estimates that are encountered in a real world data set. The choice of variables is not based on relevant applied research and thus the results should not be used to draw inferences about the student data that are being analyzed.

Data Set

The STAR database provides raw student- and classroom-level data collected for a longitudinal experiment conducted by the Tennessee Department of Education (Achilles, C.M. et al., 2008). This was a four-year longitudinal class-size study, where over 7,000 students across 79 schools were randomly assigned to one of three class-size interventions: 1) a small class (13 to 17 students), 2) regular class (22 to 25 students), or 3) regular-with-aide class (22 to 25 students, plus a teacher's aide). These interventions began in 1985 and followed cohorts of students from their kindergarten year (1985) through their third grade year (1989), adding new students every school year through the four years. Through the length of the project, several outcomes were periodically collected for the students, including student and teacher demographics, academic achievement test scores (norm-referenced and criterion-referenced), and behavioral measures (such as self-concept scores). After the completion of the STAR project, middle school and high school academic achievement was continuously collected for many of the participants by means of ancillary studies conducted statewide (in Tennessee), however, the present analysis only included the original 4-year longitudinal project and measures.

Project participation was open to all Tennessee school systems and, of those that expressed interest in participation, 79 schools in 42 districts participated. Within these participating schools, 328 kindergarten classes were randomly assigned to one of the three class-size interventions described above. Throughout the four years of the project, achievement data was available for 5,907 kindergarten students, 6,684 first graders, 6,559 second graders, and 6,464 third graders.

The present analysis employed only the data available for kindergarten and first grade students, including student and classroom level data. Students in this dataset were nested within a cross-classification of kindergarten classrooms and first grade classrooms. Both of these cross-classification factors were purely nested within elementary schools. Student who left the school during the project were no longer followed, and were therefore not included in the analysis. The final data set was comprised of 4,327 students cross-classified in 308 kindergarten classrooms and 332 first grade classrooms. These schools were nested in 75 elementary schools. Of this group, academic achievement scores (for reading and math) were available for 4,129 of the students.

The average number of students enrolled in each kindergarten class was 20 (with a standard deviation (SD) of 4). However, outcome data of interest, such as academic achievement scores, were not necessarily collected for every student. For this dataset, data were collected for 14 students, on average, for *each* kindergarten class. The average number of students enrolled in each first grade class was 21 (with a standard deviation (SD) of 4). As was seen in the kindergarten collection, outcome data was not collected for all first grade students. Data were collected for about 13 students, on average, for each first grade classroom. As previously mentioned, students were cross-classified by kindergarten class and first grade class, purely nested in elementary schools. When organized in a tabular format, like that in Table 2, specifics about cell sizes could be investigated for each elementary school. For this specific data set, cell sizes ranged from 1 to 17, meaning that in cells that contained at least one student, the number of students in a particular kindergarten and first grade classroom combination ranged from 1 student to 17 students, with the average number of students in each cell being 4.59 (SD = 2.75).

Dependent Variable

The outcome variable for Study 1 was the Stanford Achievement Test (SAT), a norm-referenced assessment developed by the Psychological Corporation (1983). The SAT is actually a set of standardized subtests, including math and reading subtests, used as assessments for students in kindergarten and first grade. Although each student received numerous subtests, the current study only investigated reading outcome scores from the reading comprehension subtests. This reading outcome was provided as a total reading scale score, which ranged from 315 to 627, for kindergarten students and 404 to 651 for first grade students. The data set analyzed in this study included SAT reading comprehension test scores for 4,129 students who attended kindergarten and first grade in the project schools.

Predictor Variables

Three predictor variables were included in the estimating models for this study: one student-level predictor, gender (X), and two level-2 predictors, kindergarten class size (W) and first grade class size (Z). Although the use of gender as a predictor was not based on any pre-existing substantive theory, it is sometimes included as a level-1 predictor and here was used to investigate any differences in estimation of its coefficient. The two level-2 predictors, *W* and *Z*, are each associated with one of the two cross-classified factors in this analysis- kindergarten and first-grade classrooms, respectively. These two predictors were included in the model since the STAR data project was actually launched to investigate class size interventions. Using class size as a predictor, therefore, made sense, however, no other pre-existing theories were used in deciding to include them. For this study, no level-3 predictors were selected. The current study was specifically conducted to provide a simple example of a three-level cross-classified structure, with cross-classification at level-2, which may be encountered in the real world.

Analysis

The conditional CCREM models for *Model A* and *Model B* were estimated, with the three predictors described above: gender at level-1 (*X*) and, at level-2, kindergarten class size (*W*) and first grade class size (*Z*). *Model A* and *Model B* both modeled students nested within a cross-classification of kindergarten class and first grade class at level-2, and purely nested within elementary school at level-3. SAS PROC MIXED was used to fit both models to the STAR data set described above. Again, the two models only differed in that the first model, *Model A*, assumed that the variance $\tau_{0j_1xj_2k}$ was equal to zero and *Model B* included the estimation of this variance in the analysis.

Model A was a three-level conditional CCREM model, with predictors at level-1 and level-2. The equation for level-1 was:

$$Y_{i(j_1,j_2)\mathrm{k}} = \pi_{0(j_1,j_2)\mathrm{k}} + \pi_{1(j_1,j_2)\mathrm{k}} X_{i(j_1,j_2)\mathrm{k}} + e_{i(j_1,j_2)\mathrm{k}}, \tag{30}$$

where $Y_{i(j_1,j_2)k}$ was the reading SAT score; $X_{i(j_1,j_2)k}$ was the gender of student *i* attending kindergarten class j_1 and first grade class j_2 in elementary school *k*; $\pi_{0(j_1,j_2)k}$ was the expected reading SAT score for students in kindergarten class j_1 and first grade class j_2 in elementary school *k* when $X_{i(j_1,j_2)k}$ is zero; and $\pi_{1(j_1,j_2)k}$ was the change in reading SAT score for every one unit change in *X*, holding all else constant. $\pi_{1(j_1,j_2)k}$ was assumed to be fixed, not randomly varying, across the level-2 (kindergarten class and first grade class) and level-3 (elementary school) factors. $e_{i(j_1,j_2)k}$ was the random effect associated with student $i(j_1, j_2)k$, with an assumed mean of zero and variance $\sigma^2_{i(j_1,j_2)k}$.

At the second level of *Model A*, the cross-classification was modeled. The equation used for level-2 is expressed as:

$$\begin{cases} \pi_{0(j_1,j_2)\mathrm{k}} = \beta_{000\mathrm{k}} + \beta_{010\mathrm{k}} W_{0(j_1,j_2)\mathrm{k}} + \beta_{001\mathrm{k}} Z_{0(j_1,j_2)\mathrm{k}} + u_{0j_10k} + u_{00j_2k} \\ \pi_{1(j_1,j_2)\mathrm{k}} = \beta_{100\mathrm{k}} \end{cases}, \tag{31}$$

where β_{000k} was the expected SAT score for school *k*, when $W_{0(j_1,j_2)k}$ and $Z_{0(j_1,j_2)k}$ are both equal to 0; u_{0j_10k} was the random effect associated with kindergarten classroom j_1 in elementary school *k*; and u_{00j_2k} was the random effect associated with first grade classroom j_2 in elementary school *k*. Both random effects were assumed to be independently and normally distributed, with a mean of zero and a variance of τ_{0j_10k} and τ_{00j_2k}, respectively. β_{010k} was the expected change in the reading outcome for every one unit change in *W*, holding all else constant; β_{001k} was the expected change in the reading outcome for every one unit change in *Z*, holding all else constant. β_{010k} and β_{001k} were assumed to be fixed, not randomly varying, across the other level-2 cross classified factor (first grade classroom), and the level-3 (elementary school) factor.

The third level of *Model 1* was purely hierarchical. The equation at level-3:

$$\begin{cases} \beta_{000k} = \gamma_{0000} + r_{000k} \\ \beta_{010\mathrm{k}} = \gamma_{0100} \\ \beta_{001\mathrm{k}} = \gamma_{0010} \\ \beta_{100\mathrm{k}} = \gamma_{1000} \end{cases}, \tag{32}$$

where γ_{0000} was the predicted reading SAT score, when *X, W,* and *Z* are each equal to zero; γ_{0100} was the change in the outcome score, with a one unit change in predictor, *W*, holding all else constant; γ_{0010} was the change in the outcome score, with a one unit change in predictor, *Z*, holding all else constant; γ_{0010} was the is the change in the outcome score, with a one unit change in predictor, *X*, holding all else constant; and r_{000k} was the random effect associated with elementary school *k*. The random effect was assumed to be normally distributed, with a mean of zero and a variance of τ_{000k}. While predictors at level-3 can be included, for this example, no predictors associated with the elementary school were included in this analysis.

These three equations can be combined into one equation:

$$\begin{aligned} &Y_{i(j_1,j_2)\mathrm{k}} = \gamma_{0000} + \gamma_{1000} X_{i(j_1,j_2)\mathrm{k}} + \gamma_{0100} W_{0(j_1,j_2)\mathrm{k}} + \gamma_{0010} Z_{0(j_1,j_2)\mathrm{k}} + \\ &r_{000k} + u_{0 j_1 0k} + u_{00 j_2 k} + e_{i(j_1,j_2)k} \end{aligned} \tag{33}$$

This combined equation for *Model A* illustrates how a student's SAT reading score was modeled as a function of the grand mean, the three predictors (and their coefficients) in the model, *X*, *W*, *Z*, and the random effects associated with the students' kindergarten classroom, first grade classroom, elementary school, and unknown individual differences.

The STAR data were also analyzed using a second model, *Model B*. This model was parameterized exactly like the model in Equation 33, with *one* additional random effect, $u_{0(j_1 x j_2)\mathrm{k}}$, that models the interaction of the cross-classified factors' random effects. The *Model B* equation, with all levels combined, is expressed as:

$$\begin{aligned} &Y_{i(j_1,j_2)\mathrm{k}} = \gamma_{0000} + \gamma_{1000} X_{i(j_1,j_2)\mathrm{k}} + \gamma_{0100} W_{0(j_1,j_2)\mathrm{k}} + \gamma_{0010} Z_{0(j_1,j_2)\mathrm{k}} + \\ &r_{000k} + u_{0\,j_1 0k} + u_{00\,j_2 k} + u_{0(j_1 x j_2)k} + e_{i(j_1,j_2)k} \end{aligned} , \qquad (34)$$

where the additional level-2 random effect, $u_{0(j_1 x j_2)\mathrm{k}}$, was assumed to have a mean of zero and a variance of $\tau_{0(j_1\ x\ j_2)k}$. The addition of this random effect means that three variance components associated with the crossed factors were estimated in *Model B*: $\tau_{0\,j_1 0k}$, $\tau_{00\,j_2 k}$, and $\tau_{0\,j_1 x j_2 k}$. Here, the variance of the random effect interaction, $\tau_{0\,j_1 x j_2 k}$, is the variance associated with the interaction of kindergarten classroom and first grade classroom. When this value is nonzero, it implies that for two students in the same first grade class, the effect of first grade classroom on reading SAT scores differs depending on their kindergarten classroom effects.

The two models were fit using SAS PROC MIXED with Restricted Maximum Likelihood (REML) estimation. Comparisons were made for the resulting parameter estimates and associated standard errors. Additionally, the default information criteria for

SAS PROC MIXED, Akaike's (1979) information criteria (AIC), corrected criterion (AICC; Hurvich & Tsai, 1989), and Bayesian information criterion (BIC, Schwarz, 1978), were also compared. The two models were fit to compare possible model selection differences that can occur when the estimation of $\tau_{0\,j_1xj_2k}$ is included in one of the models.

Results

Applied substantive conclusions should not be drawn about the STAR analysis results in this study. Table 3 and Table 4 present the parameter estimates, associated standard errors (SE), and fit indices for *Model A* and *Model B.*

Fixed Effects

Estimations for both models included four fixed effect estimates: 1) the level-1 intercept, γ_{0000}, 2) the coefficient of the level-1 predictor, γ_{1000}, 3) the coefficient of the level-2 predictor, γ_{0100}, which is associated with the first cross-classified factor, kindergarten class, and 4) the coefficient of the level-2 predictor, γ_{0010}, which is associated with the second cross-classified factor, first grade classroom.

As seen in Table 3, for all levels, estimates of the fixed effect coefficients were relatively unaffected by the assumption that the random effects' interaction variance $\tau_{0(j_1xj_2)k}$ was zero in the estimating model ***(****Model A****)***. A similar pattern is found for the level-2 coefficients of *W* and *Z*, and their associated standard errors. Differences in these estimates, and their associated standard errors, between the two models ranged from 0.00 to 0.11. Overall, only negligible differences between *Model A* and *Model B* were found for all fixed effect estimates and their associated standard errors.

Table 3

Parameter (and Standard Error) Estimate Comparisons between Model A and Model B

	Model A			*Model B*		
Parameter	Coeff.	Estimate	(SE)	Coeff.	Estimate	(SE)
Fixed Effect						
Intercept	γ_{0000}	554.18*	(7.26)	γ_{0000}	554.29*	(7.24)
X	γ_{1000}	11.97*	(1.51)	γ_{1000}	11.95*	(1.51)
W	γ_{0100}	-0.61	(0.36)	γ_{0100}	-0.62	(0.36)
Z	γ_{0010}	-1.09*	(0.40)	γ_{0010}	-1.09*	(0.40)
Random Effect Variance						
Students	$\sigma_{i(j_1,j_2)k}$	2,249.19*	(52.52)	$\sigma_{i(j_1,j_2)k}$	2,240.56*	(54.47)
Kinder	τ_{0j_10k}	17.67	(19.70)	τ_{0j_10k}	11.18	(23.14)
First	τ_{00j_2k}	222.11*	(37.97)	τ_{00j_2k}	213.95*	(40.80)
Kinder x First Class	$\tau_{0(j_1xj_2)k}$	-----	-----	$\tau_{0(j_1xj_2)k}$	21.84	(40.23)
Elementary	r_{000k}	630.50*	(122.21)	r_{000k}	629.65*	(121.87)

Note. $\tau_{0j_1xj_2k}$ was not estimated for *Model A*, but was estimated for *Model B*. Coeff. = coefficient; SE = standard error estimate.
* p < .01. ---- = not estimated.

Random Effects

Random effects' variance components and their associated standard errors were also estimated for *Models A* and *B*, and are presented in Table 3. Unlike the fixed effect estimates, estimates for the random effects variance components did differ depending on the estimating model.

For the level-1 random effect, the variance associated with students, $\sigma^2_{i(j_1 x j_2)k}$, was slightly larger for *Model A* (2,249.19) than for *Model B* (2,240.56). The associated standard error, however, was slightly smaller in *Model A* (52.52) than for *Model B* (54.47). Estimation differences were also observed for the variances associated with the cross-classification factors (here, kindergarten and first grade classrooms). For kindergarten classrooms, *Model B* resulted in a lower between kindergarten classroom variance estimate, $\tau_{0 j_1 0 k}$, (11.18) than *Model A* (17.67). The associated SE estimates for kindergarten classroom showed an opposite trend. The estimate for *Model B* (23.14) was slightly higher than for *Model A* (19.70). The same trend can be seen for the variance associated with first grade classrooms, $\tau_{00 j_2 k}$. Estimations from *Model B* were smaller (213.95) than for *Model A* (222.11). The associated SE estimate for first grade classroom were larger for *Model B* (40.80) than for *Model A* (37.97). For the elementary school variance, r_{000k}, model differences were very slight, less than 1.00 for both the variance estimation and the associated standard error.

Information Criteria

Information criteria were also provided in the results. Information criteria comparisons between *Model A* and *Model B* are listed in Table 4. Here, the three information criteria were looked at (AIC, AICC, BIC), because these are the default information criteria reported by SAS PROC MIXED. When comparing information criteria between two models (here, AIC, AICC, BIC), smaller values imply better fit. Here, only slight differences were observed between the two models. For all three information criteria, *Model A* provided slightly smaller values, all less than a difference of 4.00. This does not imply that *Model A* is the most correct model for the data, only that it is the better fitting model between the two models that were compared using this dataset.

Table 4

Information criteria for Model A and Model B

Information Criteria	*Model A*	*Model B*
AIC	44,014.8	44,016.5
AICC	44,014.8	44,016.5
BIC	44,024.0	44,028.0

Note. $\tau_{0\,j_1 x j_2 k}$ was not estimated for *Model A*, but was estimated for *Model B.*

Summary

This real data analysis was intended only to assess the potential impact of including the estimation of the random effects' interaction variance, $\tau_{0(j_1 x j_2)k}$, for data with cross-classification at level-2. The analysis of the STAR data set indicated that assuming $\tau_{0(j_1 x j_2)k}$ is zero, and not including its' estimation in the model did provide different estimates for some of the variances and their associated standard errors than when this variance was estimated. This supports the CCREM methodological findings of Shi et al. (2010), and the findings of Luo and Kwok (2009) for misspecified models with cross-classification at level-2 of a three-level model. Notably, when $u_{0(j_1 x j_2)\mathrm{k}}$ was not included in the CCREM model, the associated variance, $\tau_{0(j_1 x j_2)k}$, (which is assumed to be zero) is redistributed to other variance components. The estimates for the level-1 and level-2 variances were slightly larger when $\tau_{0(j_1 x j_2)k}$ was assumed to be zero, and not estimated. While these findings were not meant to provide evidence about the nature of the random effects' interaction, $u_{0(j_1 x j_2)\mathrm{k}}$, the differences in estimates of the variance components in *Model A* and *Model B* does warrant further investigation.

Additionally, the within-cell sample sizes for the STAR dataset seemed sufficiently large to allow for reasonable estimation of this variance component, $\tau_{0(j_1 x j_2)k}$. However, since specific cell sample size guidelines are not available, further analysis of appropriate within-cell sample sizes which allow for good recovery of the $\tau_{0(j_1, j_2)k}$ variance should be explored. The next study utilized a Monte Carlo simulation to explore the properties that might influence estimation of the variance component for the random effects' interaction for three-level CCREM models, with cross-classification at the second level. Also, some of the parameter estimates found in study 1 were used to provide realistic parameter values that can be encountered in educational data.

STUDY 2 - SIMULATION

The present simulation compared two estimating models, one model included the variance component, $\tau_{0(j_1, j_2)k}$, and the second model assumed the value of this variance component was zero, under several manipulated conditions. The simulation allowed for investigation of these manipulated conditions and their impact on parameter recover. The present simulation also investigated estimation issues that may occur for varying cell sizes.

Generating Conditions

In Study 2, data were generated to follow the three-level cross-classified structure described in Luo and Kwok (2009) and found in the STAR dataset, with students cross-classified by kindergarten classroom (F1) and first grade classroom (F2), purely nested in elementary schools, with a continuous student level outcome. In this study several conditions were manipulated for a three-level CCREM, and their impact on parameter recovery was investigated. The conditions included: classroom sample size (small, average, large), the structure of the cross-classification, the intra-unit correlation coefficient (IUCC), (7% and 13%), and the cross-classified factors' variance component values (equal and unequal). Findings for the STAR data set in Study 1 were used as guidelines for parameter values in conditions where applicable. Table 5 presents all of the conditions investigated in this study.

Mean Classroom Sample Size

For Study 2, the number of level-1 units, here, students, were selected from three classroom sample sizes, representing classroom sizes that were found in the STAR data from the previous analysis. In the STAR data set, the mean classroom sample size for kindergarten classrooms was 20 students, with a standard deviation of approximately 4.

The first classroom sample size is considered small, defined as 16 students. This number was considered representative of a small size, because it is selected as one standard deviation below the mean. The second classroom sample size was the "average" class size, here defined as 20 students, which was the mean number of students in a kindergarten classroom for the STAR data. Finally, the large classroom sample size was defined as 24 students, which was one standard deviations above the mean classroom size in the STAR data. Values for these three sample sizes were selected to contribute to specific within-cell sample sizes that might be considered small for appropriate estimations of (Raudenbush and Bryk, 2002).

Table 5

Conditions for Monte Carlo Simulation Study

Condition	Classroom sample size	Partial cross-classification structure	IUCC	Level-2 variance components
1	small	2 feeder	7%	Equal
2	small	2 feeder	7%	Unequal
3	small	2 feeder	13%	Equal
4	small	2 feeder	13%	Unequal
5	small	4 feeder	7%	Equal
6	small	4 feeder	7%	Unequal
7	small	4 feeder	13%	Equal
8	small	4 feeder	13%	Unequal
9	average	2 feeder	7%	Equal
10	average	2 feeder	7%	Unequal
11	average	2 feeder	13%	Equal
12	average	2 feeder	13%	Unequal
13	average	4 feeder	7%	Equal
14	average	4 feeder	7%	Unequal
15	average	4 feeder	13%	Equal
16	average	4 feeder	13%	Unequal
17	large	2 feeder	7%	Equal
18	large	2 feeder	7%	Unequal
19	large	2 feeder	13%	Equal
20	large	2 feeder	13%	Unequal
21	large	4 feeder	7%	Equal
22	large	4 feeder	7%	Unequal
23	large	4 feeder	13%	Equal
24	large	4 feeder	13%	Unequal

Note. IUCC = Intra-unit correlation coefficient

Cross-classified Data Structures

Luo and Kwok (2009) manipulated the structure of the cross classification by varying the number of feeder and receiver classification factors. By manipulating the number of feeders/receivers, partial cross-classification became more complete as the number of feeders/receivers increased. For the present simulation, the first cross-classified factor, kindergarten classroom (F1), is called the feeder. That is, the level-1 units, here, students, *feed* into the second cross-classified factor, first grade classroom (F2) from F1. Luo and Kwok defined *complete* cross-classification as occurring in a dataset when the probability of a lower level unit affiliating with *any* cluster of the upper level factor as being approximately equal across all units. Specifically, a level-1 unit, student, from one kindergarten classroom has an equal probability of attending any first grade classroom in any school. *Partial* cross-classification occurs when the level-1 unit, here, student, from one kindergarten classroom can only attend only some – not all – of the first grade classrooms. In some educational settings partial cross-classification is considered more realistic because students are often nested in schools, or classrooms, that follow a specific feeder pattern, thereby making the equal probability of attending any combinations pattern less realistic (Luo & Kwok, 2009). The present study only investigated conditions of partial cross-classification, however, by increasing the number of feeders in a given condition, the data structure will become more cross-classified.

Investigation of the STAR data set showed that several patterns of partial cross-classification were present. In cases where there was less partial cross-classification, students from a particular kindergarten classroom were only associated with *some* of the first grade classrooms in the same school. A more complete cross-classification structure was observed as students from a particular kindergarten class were associated with every first grade classroom within that elementary school. The cross-classification structure can be thought of as a range that goes from partial to complete cross-classification. For the STAR data, as the number of students from a given kindergarten classroom are associated with more first grade classrooms, the data became closer to complete cross-classification. The current study utilized two different feeder patterns to imitate two patterns of cross-classification that went from partial to more complete cross-classification. A fully complete cross-classification structure is not realistic for this example, and was therefore not investigated.

Increasing the number of feeders increased the cross-classification structure in that more cells in the data structure contain units. Following Luo and Kwok's method of creating partial cross-classification, which is described in detail in the next section, all 4 first grade classrooms in a school were selected as receivers, with either 2 or 4 kindergarten classrooms as feeders (condition-dependent). The condition with 2 feeders corresponded with a partial cross-classified structure, where the condition with a larger number of feeders (4 feeders) corresponded with a more complete cross-classified data structure.

IUCC

The IUCC functions similarly to the ICC in that both coefficients can be used to estimate the proportion of total variance attributed to specific levels and variables.

Applying an unconditional model to the STAR dataset, IUCC values were calculated. These calculations found that the IUCC values between classrooms for first grade classrooms was 0.07, 0.01 for kindergarten classrooms, 0.01 for the first grade by kindergarten classroom interaction, 0.20 for elementary schools, and 0.72 for the level-1 residuals variance (among students). That is, the proportion of variability attributed to the first grade classroom attended, was 7%. Additionally, the variability attributed to kindergarten classroom was 1%, the variability attributed to the first grade by kindergarten classroom interaction was 1%, the variability attributed to the elementary school attended was 20%, and to individual differences was 72%.

In their examination of group-randomized trials, Spybrook and Raudenbush (2009) investigated several educational research studies funded by the Institute of Education Sciences. When reviewing design parameters for studies that utilized three-level randomized trials (students nested in clusters, nested in sites), they found that ICCs for level-2 tended to be smaller than those for level-3 units, even after accounting for demographic covariates. Specifically, they found that when variance was partitioned into three factors (student, classroom, and school) the range for classroom ICCs was 0.07 to 0.13 for educational outcomes. That is, classroom was contributing to between 7% and 13% of the variance in educational outcomes. The lower value of this range corresponds to the first grade IUCCs found in our STAR data set, 7%. This provided two values for the IUCC condition that was manipulated in this simulation study: 7% and 13%. These IUCC values (7% and 13%) are used to manipulate the value of the variance components as described below.

Equality of the Level-2 Variance Components

As seen in Equation 26, the variance components for the random effects, u_{0j_10k}, u_{00j_2k}, and $u_{0(j_1\ X\ j_2)\mathrm{k}}$, are assumed to be independently normally distributed with means of zero and variances of τ_{0j_10k}, τ_{00j_2k}, and $\tau_{0(j_1\ X\ j_2)k}$, respectively. When investigating CCREM, τ_{0j_10k}, τ_{00j_2k}, and $\tau_{0(j_1\ X\ j_2)k}$ have typically been generated to be equal in value (Meyers and Beretvas, 2006; Shi et al., 2010). The current study manipulated two conditions for the variance components, τ_{0j_10k}, τ_{00j_2k}, and $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. In the first condition, values for these variance components were generated to be equal, where $\tau_{0(j_1,j_2)k} = \tau_{0j_10k} = \tau_{00j_2k}$. Here, the value of the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, is generated to be equal to the values of the two level-2 cross-classification variances, and the two level-2 cross-classification variances were generated to be equal to each other. That is, all of the level-2 variance components were generated to be equal. In the other set of conditions, the pattern for variance component values were generated to be unequal such that: $\tau_{0(j_1xj_2)k} \neq \tau_{0j_10k}$ or τ_{00j_2k}. Here, the value of the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, is generated to be less than (unequal) the values of the remaining level-2 cross-classification variances. This provides one set of conditions where $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is the exact same value as the remaining level-2 variance components, and another set where it is *less than* the remaining level-2 variance components.

Generation and Estimation

One thousand datasets for every combination of conditions was generated using SAS software (version 9.3). The following sections provide detailed explanations of model generation and estimation.

Generating Model

One thousand datasets per condition were generated to fit a conditional three-level cross-classified model, with students at level-1, cross-classification of kindergarten classrooms and first grade classrooms at level-2, and pure nesting in elementary schools at level-3. This data structure followed the one found in the STAR data set of study 1. In conditions where model estimation did not converge (i.e. one variance component or coefficient was not estimated) additional replications were generated until the number of converged solutions reached 1,000.

Predictors were included in the current simulation, mimicking the predictors used in the analysis for Study 1. Gender was used as the student-level predictor, *X*, or the level-1 predictor, and kindergarten class size and first grade class size were used as the predictors for level-2, *W* and *Z*, respectively. Each of these level-2 predictors is associated with one of the crossed factors. The generating model equation was as follows:

$$Y_{i(j_1,j_2)k} = \pi_{0(j_1,j_2)k} + \pi_{1(j_1,j_2)k} X_{i(j_1,j_2)k} e_{i(j_1,j_2)k}, \tag{35}$$

for level-1,

$$\begin{cases} \pi_{0(j_1,j_2)\mathrm{k}} = \beta_{000k} + \beta_{010k} W_{0(j_1,j_2)\mathrm{k}} + \beta_{001k} Z_{0(j_1,j_2)\mathrm{k}} + u_{0j_10k} + u_{00j_2k} + u_{0(j_1xj_2)k} \\ \pi_{1(j_1,j_2)\mathrm{k}} = \beta_{100k} \end{cases} \tag{36}$$

for level-2, and

$$\begin{cases} \beta_{000k} = \gamma_{0000} + r_{000k} \\ \beta_{010k} = \gamma_{0100} \\ \beta_{001k} = \gamma_{0010} \\ \beta_{100k} = \gamma_{1000} \end{cases} \tag{37}$$

for level-3, with a continuous, student level outcome, cross-classification at level-2, and variances u_{0j_10k}, u_{00j_2k}, and $u_{0(j_1xj_2)k}$ sampled from independent normal distributions each with a mean of zero and with condition specific variances of , τ_{0j_10k}, τ_{00j_2k}, and $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

Together the generating model can be combined as:

$$Y_{i(j_1,j_2)k} = \gamma_{0000} + \gamma_{1000}(X) + \gamma_{0100}(W) + \gamma_{0010}(Z) + r_{000k} + u_{0j_10k} + u_{00j_2k} + u_{0(j_1xj_2)k} + e_{i(j_1,j_2)k} \tag{38}$$

Generating Fixed Effects

Once again using the values from the STAR data set as a guide, three fixed effect were generated for the model in Equation 38. The fixed effect representing gender, *X*, was generated, so that 50% of the students in the data set were randomly assigned a 0 for male and the other 50% a 1 for female. This mimicked the approximate ratio of male to females found in the STAR dataset. The fixed effects representing overall classroom size for kinder and first, *W* and *Z* respectively, were randomly generated for each classroom by selecting a value from a normal distribution of 20, with a standard deviation of 4. The coefficients for the three predictors, as well as the intercept, were also generated from the values found in the STAR data set. Generating values used for the parameters, γ_{0000}, γ_{1000}, γ_{0100}, and γ_{0010} were 554, 11.95, -1, and -1, respectively. Contextual effects were not investigated in this study and, therefore, the simulated predictors were not grand- nor group-mean centered.

Variance Component Values

Equal level-2 variances conditions. For this condition, values were calculated that corresponded to IUCC values of 7% and 13%, condition dependent. For both IUCC values the total variance found in Study 1 was used as a starting reference. This value of 3,117 was set to be the total variance value to begin calculations of the other variance components, using formulas found in Equations 27, 28, and 29. The other variance component values were then set accordingly, depending on the IUCC value.

For conditions where the IUCC value was 7%, 7% of 3,117 was used as the value for the variances associated with kindergarten classroom, first grade classroom, and the interaction of both. For elementary school variance, 20% of 3117 was used as the value for the variance of elementary school. This corresponds to the 20% IUCC for elementary school that was found for the STAR data set. The remaining variance for 3,117 was then used as the value for the level-1 residual.

For conditions where the IUCC value was 13%, 13% of 3,117 was used as the value for the variances associated with kindergarten classroom, first grade classroom, and the interaction of both. For elementary school variance, 20% of 3,117 was used as the value for the variance of elementary school. This corresponds to the 20% IUCC for elementary school that was found for the STAR data set. The remaining variance for 3,117 was then used as the value for the level-1 residual.

Unequal level-2 variances conditions. For this condition, values were calculated that corresponded to IUCC values of 7% and 13%, condition dependent. Again, for both IUCC values the total variance found in Study 1 was used as a starting reference. This value of 3,117, was set to be the total variance value to begin calculations of the other variance components, using formulas found in Equations 27, 28, and 29. The other variance component values were then set accordingly, depending on the IUCC value.

For conditions where the IUCC value was 7%, 7% of 3,117 was used as the value for the variances associated with kindergarten classroom and first grade classroom. For this condition, the level-2 variance, $\tau_{0(j_1 x j_2)k}$ was generated to be smaller than the variances for kindergarten classroom and first grade classroom. The new value was calculated as 3.5% of 3,117, which is half of the 7% used for the variances of kindergarten and first grade classroom. For elementary school variance, 20% of 3,117 was still used as the value for the variance of elementary school. The remaining variance for 3,117 was then used as the value for the level-1 residual.

For conditions where the IUCC value was 13%, 13% of 3,117 was used as the value for the variances associated with kindergarten classroom and first grade classroom. Again, for this condition, the level-2 variance, $\tau_{0(j_1 x j_2)k}$ was generated to be smaller than the variances for kindergarten classroom and first grade classroom. The new value was calculated as 6.5% of 3,117, which is half of the 13% used for the variances of kindergarten and first grade classroom. For elementary school variance, 20% of 3,117 was still used as the value for the variance of elementary school. The remaining variance for 3,117 was then used as the value for the level-1 residual.

Generating Cross-classification

Following the cross-classification procedure presented in Luo and Kwok (2009) and using the mean number of kindergarten and first grade classrooms from the STAR data set as a guide, for all conditions 4 kindergarten classrooms were generated to be purely nested in 4 first grade classrooms, purely nested in 75 elementary schools. Table 6 presents an example of how the kindergarten classrooms feed into the first grade classrooms for a large classroom sample size in one elementary school.

Table 6

Generating Cross-classification

		First grade class			
Feeder	**Kindergarten class**	*A*	*B*	*C*	*D*
1 feeder	*A*	24			
	B		24		
	C			24	
	D				24
2 feeder	*A*	12			12
	B	12	12		
	C		12	12	
	D			12	12
4 feeder	*A*	12	4	4	4
	B	4	12	4	4
	C	4	4	12	4
	D	4	4	4	12

The purely nested structure, corresponding to a feeder value of 1, was the *starting* point of the data generation, where here, students in these kindergarten classrooms can be considered to be attending their designated first grade classroom. All twenty-four students from kindergarten class A were sent to first grade class A. To begin the cross-classification process, 50% of the students would stay and attend their "designated" first grade classroom. The kindergarten class would then send the other 50% to either one non-designated first grade classroom (2 feeder) or the remaining first grade classrooms (4 feeder), within the same elementary school. This is done for all 4 kindergarten classes in a given elementary school.

Analysis

SAS PROC MIXED was used to fit two estimating models. Restricted Maximum likelihood (REML) was used as the estimation method. The first model, *Model 1* corresponds to the model in Equation 38, where students at level-1 are cross-classified by kindergarten and first grade classrooms at level-2, nested in elementary schools at level-3. Three predictors were included in the model, one level-1 predictor, *X*, and two level-2 predictors, *W* and *Z*, similar to the analysis in Study 1. *Model 1* also included estimation of the variance of the random effects interaction, $u_{0(j_1 x j_2)k}$. The equation for *Model 1* was:

$$Y_{i(j_1,j_2)k} = \gamma_{0000} + \gamma_{1000}(X) + \gamma_{0100}(W) + \gamma_{0010}(Z) + r_{000k} + u_{0j_10k} + u_{00j_2k} + u_{0(j_1 x j_2)k} + e_{i(j_1,j_2)k}. \quad (39)$$

To provide a comparison with the model typically estimated in applied research, the present simulation also included a second estimating model that did not estimate the variance $\tau_{0(j_1 x j_2)k}$. *Model 2* was similar to *Model 1*, but **did not** include the random effects' interaction, $u_{0(j_1 x j_2)\mathrm{k}}$:

$$Y_{i(j_1,j_2)k} = \gamma_{0000} + \gamma_{1000}(X) + \gamma_{0100}(W) + \gamma_{0010}(Z) + r_{000k} + u_{0j_1 0k} + u_{00j_2 k} + e_{i(j_1,j_2)k}. \quad (40)$$

For both models, parameter estimates and associated standard errors were assessed using relative parameter bias and standard error bias.

Relative Parameter Bias

Relative parameter bias was calculated for the each fixed effect and random effect variance component for both of the estimating models. Relative parameter bias was calculated using the formula:

$$RPB\left(\hat{\theta}\right) = \frac{\bar{\hat{\theta}} - \theta}{\theta}, \quad (41)$$

where $\bar{\hat{\theta}}$ represents the average parameter estimate across the 1,000 replications that were conducted for the study (Hoogland & Boomsma, 1998). In order to assess if any substantial bias is found in the parameter of interest, the cutoff criteria specified in Hoogland and Boomsma (1998) was utilized, where an absolute value of $RPB(\hat{\theta})$ less than 0.05 is considered to be an acceptable amount of bias.

Standard Error Bias

In addition to relative parameter bias, relative standard error bias was also calculated for the standard error estimates associated with the fixed and random effects. The formula used for calculating the relative bias of the standard error was:

$$RSEB\left(S_{\hat{\theta}}\right) = \frac{\bar{S}_{\hat{\theta}} - SD(\hat{\theta})}{SD(\hat{\theta})}, \quad (42)$$

where, $\bar{S}_{\hat{\theta}}$ is the average of the standard error estimates calculated across the 1,000 replications and $SD(\hat{\theta})$ is the empirical standard error. The empirical standard error is obtained by calculating the standard deviation of the parameter estimates obtained across all 1,000 replications for each condition. Hoogland and Boomsma's (1998) cut-off criteria was also utilized here, where an acceptable level of bias is a value of $RSEB(\hat{\theta})$ that is less than .10.

Analysis of Variance

In addition to the bias estimates described above, analyses of variance (ANOVA) were conducted to investigate which specific conditions in the simulation design may have affected the relative parameter bias and standard error bias. ANOVAs were conducted with the relative parameter or standard error bias used as the dependent variable and each manipulated variable (or conditions) as the independent variables, or factors. For example, IUCC values were manipulated in the study, therefore IUCC was considered an independent variable in the ANOVA, and its effect on bias was calculated. ANOVAs were only conducted when substantial bias (as defined above) was found for multiple conditions. When substantial bias was encountered, a four-way between groups ANOVA was conducted for the results for each model (*Model 1* and *Model 2*) where the bias was estimated. Main effects and two-way interactions were investigated for any of the ANOVAs that were conducted.

All ANOVAs provided tests of statistical significance, however, the large sample size included in these ANOVAs may have provided too much power, where even the smallest true effect would have been statistically significant. Therefore, practical significance rather than statistical significance was interpreted. Partial eta squared (η_p^2) values were calculated to provide a measure of practical significance. Following previous research, (Kirk, 1995), a cutoff of 0.01 or greater was used to determine practical significance. That is, any main effect or interaction with a partial eta squared value of 0.01 or larger was considered practically significant.

Information Criteria

Lastly, to identify the best fitting model, the SAS PROC MIXED default indices; Akaike's Information Criterion (AIC), AICC, and BIC were obtained for both *Model 1* and *Model 2*. Because data was generated to include the random effects interaction, $u_{0(j_1xj_2)k}$, corresponding to *Model 1*, a tally was kept for the proportion of each set of 1,000 iterations in each condition where the better fit of the correct model (*Model 1*) was chosen over the fit of *Model 2*. For all three criterion values, a smaller value indicates a better fitting model.

Chapter 4: Results

This chapter summarizes the results of the Monte Carlo simulation conducted, where two different CCREM models were used to estimate 1,000 datasets for each manipulated condition. The first model, *Model 1*, included estimation of the variance of the random effects' interaction, $\tau_{0(j_1,j_2)k}$, and in the second model, *Model 2*, this variance component was assumed to equal zero. First, findings for the relative parameter bias are described for the fixed effects estimates and the random effects' estimates. Then, findings for the relative standard error (*SE*) bias are presented for both fixed and random effects' estimates. Lastly, a description of the information criteria results are also presented for all 24 conditions.

RELATIVE PARAMETER BIAS

Fixed Effects

Relative parameter bias was calculated for four fixed effect estimates: the coefficient of the student level predictor, γ_{1000}, the coefficients for the two level-2 predictors associated with the crossed factors, γ_{0100} and γ_{0010} , and the intercept at level-1, γ_{0000}. These bias values were estimated for *Model 1*, the correct model that included the estimation of $\tau_{0(j_1,j_2)k}$ (which had been generated to be non-zero) and *Model 2*, the misspecified model, which involved an assumption that $\tau_{0(j_1,j_2)k}$ was zero and is most commonly used by applied researchers estimating a CCREM. Parameter estimates were considered to be substantially biased if the absolute value of the relative parameter bias was greater than 0.05 (Hoogland and Boomsma, 1998). Again, ANOVAs were conducted when substantial parameter bias was found for multiple conditions. Main effects and two-way interactions were investigated and partial eta squared (η_p^2) values larger than 0.01 were considered practically significant.

Intercept, γ_{0000}

Relative parameter bias for estimates of the intercept, γ_{0000}, are presented in Table 7. This table summarizes bias for both the *Model 1* and *Model 2* estimates in each of the 24 conditions. No substantial parameter bias was found for any of the conditions, therefore an ANOVA was not conducted. Across all conditions, relative bias remained nearly the same, with values of approximately 0.001, well below the acceptable cutoff of 0.05 (Hoogland & Boomsma, 1998). Within the same condition, relative bias for the estimates of γ_{0000} were almost identical for *Model 1* and *Model 2*. Overall, parameter estimates for the intercept, γ_{0000}, were not substantially affected by the estimating model, class size, partial cross-classification structure, the cross-classified (CC) factors' variance values, or the IUCC values.

Student Level Predictor Coefficient, γ_{1000}

Relative parameter bias for estimates of the coefficient of the level-1 predictor, γ_{1000}, are presented in Table 8. There was no substantial relative parameter bias found for any of the conditions and, again, an ANOVA was not conducted. Across all conditions, relative bias for the coefficient, γ_{1000}, was similar, all with absolute values less than 0.01.

Within the same condition, relative bias for these estimates were almost identical between *Model 1* and *Model 2*. Overall, parameter estimates for the coefficient, γ_{1000}, were not substantially affected by the estimating model or any of the condition specifications (class sample size, cross-classification, etc.).

Table 7

Relative Parameter Bias for the Intercept, γ_{0000}*, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.001	0.001
			13%	0.001	0.001
		Unequal	7%	0.001	0.001
			13%	0.001	0.001
	4 feeder	Equal	7%	0.001	0.001
			13%	0.001	0.001
		Unequal	7%	0.001	0.001
			13%	0.001	0.001
Average	2 feeder	Equal	7%	0.001	0.001
			13%	0.002	0.002
		Unequal	7%	0.001	0.001
			13%	0.002	0.001
	4 feeder	Equal	7%	0.000	0.000
			13%	0.001	0.001
		Unequal	7%	0.000	0.000
			13%	0.001	0.001
Large	2 feeder	Equal	7%	0.001	0.001
			13%	0.001	0.001
		Unequal	7%	0.001	0.001
			13%	0.001	0.001
	4 feeder	Equal	7%	0.001	0.001
			13%	0.001	0.001
		Unequal	7%	0.001	0.001
			13%	0.001	0.001

Note. The $\tau_{0 j_1 x j_2 k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 8

Relative Parameter Bias for the X coefficient, γ_{1000}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.001	0.002
			13%	0.001	0.001
		Unequal	7%	0.002	0.002
			13%	0.001	0.002
	4 feeder	Equal	7%	-0.002	-0.002
			13%	-0.001	-0.001
		Unequal	7%	-0.002	-0.002
			13%	-0.002	-0.002
Average	2 feeder	Equal	7%	-0.009	-0.008
			13%	-0.007	-0.008
		Unequal	7%	-0.009	-0.008
			13%	-0.008	-0.008
	4 feeder	Equal	7%	0.000	0.000
			13%	0.000	0.000
		Unequal	7%	-0.001	-0.001
			13%	0.000	0.000
Large	2 feeder	Equal	7%	-0.002	-0.002
			13%	-0.002	-0.002
		Unequal	7%	-0.002	-0.002
			13%	-0.002	-0.002
	4 feeder	Equal	7%	0.004	0.004
			13%	0.004	0.004
		Unequal	7%	0.004	0.004
			13%	0.004	0.004

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

W Coefficient, γ_{0100}

The relative parameter bias for estimates of the coefficient of the level-2 predictor, γ_{0100}, is presented in Table 9. This is the coefficient associated with the first cross-classified factor, here, F1. There was no substantial parameter bias found for any of the study conditions and, therefore, an ANOVA was not conducted. Relative bias for the coefficient, γ_{0100}, were almost identical between *Model 1* and *Model 2*, within the same condition. Across all of the study conditions, relative bias estimates were similar, with a range of 0.001 to 0.014, below the substantial bias cut-off of .05 (Hoogland & Boomsma, 1998). Overall, parameter estimates for the coefficient γ_{0100} were not substantially affected by the estimating model or any of the study condition specifications.

Z Coefficient, γ_{0010}

Relative parameter bias for the coefficient of the second level-2 predictor, γ_{0010}, are presented in Table 10. This predictor is associated with the second cross-classified factor, here, F2. There was no substantial parameter bias was found for any of the study conditions and, therefore, an ANOVA was not conducted. Similar to the findings for the coefficient, γ_{0100}, relative bias estimates for the coefficient, γ_{0010}, were almost identical between *Model 1* and *Model 2*, within the same condition. Across all study conditions, relative bias was similar, with values ranging from -0.001 to 0.015, well below the substantial bias cut-off of 0.05 (Hoogland & Boomsma, 1998). Overall, parameter estimates for the coefficient γ_{0010} were not substantially affected by the estimating model type or any of the condition specifications.

Table 9

Relative Parameter Bias for the W coefficient, γ_{0100}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.003	0.003
			13%	0.005	0.005
		Unequal	7%	0.004	0.003
			13%	0.005	0.005
	4 feeder	Equal	7%	0.002	0.002
			13%	0.005	0.004
		Unequal	7%	0.001	0.001
			13%	0.003	0.003
Average	2 feeder	Equal	7%	0.012	0.013
			13%	0.013	0.014
		Unequal	7%	0.013	0.013
			13%	0.014	0.014
	4 feeder	Equal	7%	0.010	0.009
			13%	0.011	0.010
		Unequal	7%	0.009	0.009
			13%	0.010	0.010
Large	2 feeder	Equal	7%	0.011	0.012
			13%	0.012	0.013
		Unequal	7%	0.012	0.012
			13%	0.013	0.013
	4 feeder	Equal	7%	0.001	0.001
			13%	0.003	0.003
		Unequal	7%	0.000	0.000
			13%	0.002	0.002

Note. The $\tau_{0\,j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 10

Relative Parameter Bias for the Z coefficient, γ_{0010} *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.015	0.014
			13%	0.016	0.015
		Unequal	7%	0.015	0.014
			13%	0.015	0.015
	4 feeder	Equal	7%	0.006	0.007
			13%	0.007	0.008
		Unequal	7%	0.007	0.008
			13%	0.008	0.009
Average	2 feeder	Equal	7%	0.013	0.013
			13%	0.015	0.014
		Unequal	7%	0.013	0.012
			13%	0.014	0.013
	4 feeder	Equal	7%	-0.006	-0.006
			13%	-0.004	-0.003
		Unequal	7%	-0.006	-0.005
			13%	-0.003	0.002
Large	2 feeder	Equal	7%	0.002	0.000
			13%	0.005	0.003
		Unequal	7%	0.000	-0.001
			13%	0.003	0.001
	4 feeder	Equal	7%	0.008	0.008
			13%	0.008	0.008
		Unequal	7%	0.009	0.009
			13%	0.010	0.010

Note. The $\tau_{0 j_1 x j_2 k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Random Effects

Relative parameter bias was calculated for five random effect variance component estimates: the level-1 variance component, $\sigma^2_{i(j_1, j_2)k}$, the two cross-classified factors' variance values, $\tau_{0 j_1 0 k}$ and $\tau_{00 j_2 k}$, the level-2 variance component associated with the interaction of the crossed factors' random effects, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, and the level-3 variance component, τ_{000k}. Bias values were estimated for *Model 1*, the correct model that included the estimation of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ (which was generated to be non-zero) and *Model 2*, the misspecified model which involved the assumption that $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was zero. Again, parameters were considered to be substantially biased if the absolute value of the relative parameter bias was greater than 0.05 (Hoogland & Boomsma, 1998). When substantial bias was encountered in multiple conditions, a four-way ANOVA was conducted, which included the four condition specifications as the independent variables and the relative parameter bias of interest as the dependent variable. These ANOVAs were conducted for the model estimates where the bias occurred (either *Model 1* or *Model 2*). All main effects and two-way interactions were investigated, and partial eta-squared (η_p^2) was calculated for all factors.

Level-1 Variance, $\sigma^2_{i(j_1,j_2)k}$

Relative parameter bias for estimates of the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$ are presented in Table 11. For *Model 1*, across all conditions, no substantial bias was found and bias estimates remained at approximately 0.000. Substantial positive bias was found for 12 of the 24 conditions in *Model 2*, however, where the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was not estimated. For *Model 2*, positive parameter bias ranged from 0.015 to 0.182, indicating that the parameter, $\sigma^2_{i(j_1,j_2)k}$, was overestimated by 1.5% to 18.2%. Using a practical significance cutoff value of 0.01 for the η_p^2, results for the ANOVA indicated that three main effects (cross-classification structure, cross-classified factors' variance values, and the IUCC) and four two-way interactions affected the severity of the bias of the estimates of the level-1 variance, $\sigma^2_{i(j_1,j_2)k}$.

Table 11

Relative Parameter Bias for the Level-1 Variance Component, $\sigma^2_{i(j_1,j_2)k}$ *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.000	0.033
			13%	0.000	**0.086**
		Unequal	7%	0.000	0.016
			13%	0.000	0.037
	4 feeder	Equal	7%	0.000	**0.068**
			13%	0.000	**0.182**
		Unequal	7%	-0.001	0.032
			13%	0.000	**0.079**
Average	2 feeder	Equal	7%	0.000	0.032
			13%	0.000	**0.085**
		Unequal	7%	0.000	0.015
			13%	0.000	0.037
	4 feeder	Equal	7%	0.000	**0.058**
			13%	0.000	**0.155**
		Unequal	7%	-0.001	0.027
			13%	0.000	**0.067**
Large	2 feeder	Equal	7%	0.000	0.031
			13%	0.000	**0.083**
		Unequal	7%	0.000	0.015
			13%	0.000	0.036
	4 feeder	Equal	7%	-0.001	**0.061**
			13%	0.000	**0.163**
		Unequal	7%	-0.001	0.028
			13%	-0.001	**0.070**

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

First, the cross-classification structure of the data, 2 feeders or 4 feeders, affected the bias of the estimates of the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$ [F(1, 23,976) = 22,529.24, p < .001, η^2_p = 0.484]. Overall, the average bias of the level-1 variance estimates was lower for the 2 feeder conditions (M = 0.042) than for the 4 feeder conditions (M = 0.083), when $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ was not estimated in the model (*Model 2*). On average, across all conditions, estimates of $\sigma^2_{i(j_1,j_2)k}$ were more positively biased when the data structure was more cross-classified (and students were sent to 4 classrooms) than when the data structure was less cross-classified (and students were sent to 2 classrooms).

The cross-classified factors' variance component conditions, all equal or unequal, also affected the bias of the estimates of the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$ [F(1, 23,976) = 32,061.43, p < .001, η^2_p = 0.572]. Overall, the average bias of the level-1 variance estimates was lower when the level-2 variance component $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ was generated to be smaller than the cross-classified factors' variance values, τ_{0j_10k} and τ_{00j_2k}, (M = 0.038) than when all three cross-classified factors' variance values were generated to be equal (M = 0.086). That is, less bias was found in the level-1 variance estimates when the level-2 variance component $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ was assumed to be zero in the estimating model, and was generated to be smaller than the cross-classified factors' variance values.

The IUCC value, 7% or 13%, also affected the bias of the estimates of the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$ [F(1, 23,976) = 42,658.43, p < .001, η^2_p = 0.640]. Overall, the average bias of the level-1 variance estimates was smaller for conditions where the IUCC value was 7% (*M* = 0.035) than when the IUCC value was 13% (*M* = 0.090). On average, across all other conditions, estimates of $\sigma^2_{i(j_1,j_2)k}$ were less biased when the IUCC for the cross-classified factors, F1 and F2, were equal to 7% and increased in bias as the IUCC increased to 13%.

Although, some of the factors in the ANOVA were found to affect the severity of the bias of the $\sigma^2_{i(j_1,j_2)k}$ estimates, further analysis of the two-way interactions indicated that the effects of these factors depend on other factors in the ANOVA.

The interaction of cross-classification structure and classroom sample size affected the bias of the $\sigma^2_{i(j_1,j_2)k}$ estimates, [F(1, 23,976) = 197.40, $p < .001$, η^2_p = 0.016]. For smaller classroom sample sizes, the average bias was lower for 2 feeders (*M* = 0.043) than for 4 feeders (*M* = 0.090), resulting in a mean bias difference of 0.047. For average classroom sample sizes, the average bias was lower for 2 feeders (*M* = 0.042) than for 4 feeders (*M* = 0.077), resulting in a mean bias difference of 0.035. For large classroom sample sizes, the average bias was lower for 2 feeders (*M* = 0.041) than for 4 feeders (*M* = 0.081), resulting in a mean bias difference of 0.040. This interaction is presented in Figure 3. The differences in mean bias between the cross-classification structure conditions was largest when the classroom size was the smallest. The largest mean bias value was found for small classroom sizes with a more cross-classified structure (4 feeders).

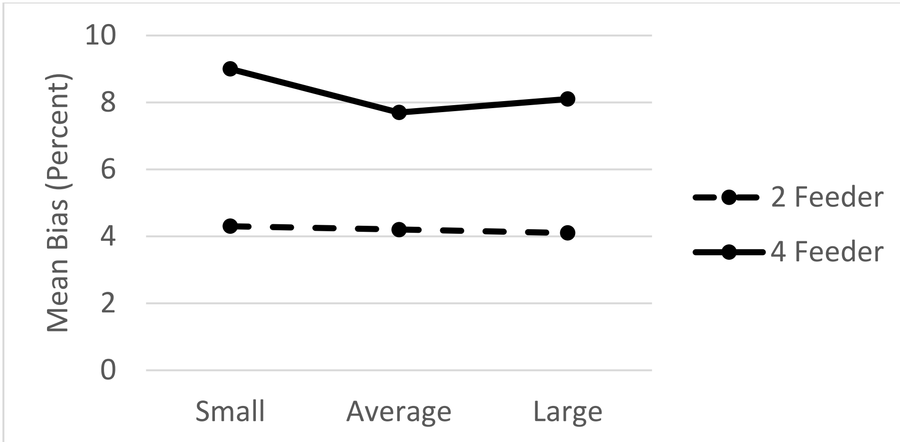

Figure 3. Cross-classification Structure and Classroom Sample Size Interaction Effect on the Mean Bias Percentage of $\sigma^2_{i(j_1,j_2)k}$.

The interaction of cross-classification structure and CC factors' variance component values affected the bias of the $\sigma^2_{i(j_1,j_2)k}$ estimates, [F(1, 23,976) = 3,419.78, $p <$ .001, η^2_p = 0.13] from *Model 2*. For conditions of 2 feeders, the average bias was lower for unequal values (M = 0.026) than for equal values (M = 0.058), resulting in a mean bias difference of 0.032. For conditions of 4 feeders, the average bias was lower for unequal values (M = 0.051) than for equal values (M = 0.115), resulting in a mean bias difference of 0.064. The difference in mean bias between the two variance value conditions was largest when the students were sent to 4 classrooms. This interaction is presented in Figure 4. Overall, as the cross-classification structure became more complete the mean bias difference between the two CC factors' variance component values also increased, from 0.032 to 0.064, a 0.032 difference.

Figure 4. Cross-classification Structure and Variance Values Interaction Effect on the Mean Bias Percentage of $\sigma^2_{i(j_1,j_2)k}$.

For *Model 2*, the interaction of cross-classification structure and IUCC values also affected the bias of the estimates of $\sigma^2_{i(j_1,j_2)k}$ [F(1, 23,976) = 4,636.57, p < .001, η^2_p = 0.162]. For conditions of 2 feeders, the average bias was lower for IUCC values of 7% (*M* = 0.023) than for 13% (*M* = 0.061), resulting in a mean bias difference of 0.038. For conditions of 4 feeders, the average bias was also lower for IUCC values of 7% (*M* = 0.046) than for 13% (*M* = 0.120), resulting in a mean bias difference of 0.074. This interaction is presented in Figure 5. The difference in average bias estimates between the two IUCC values was largest when the students were sent to 4 classrooms. Overall, as the data structure became more cross-classified the mean bias difference between the two IUCC values increased, from 0.038 to 0.074, a 0.036 difference.

Figure 5. Cross-classification Structure and IUCC Values Interaction Effect on the Mean Bias Percentage of $\sigma^2_{i(j_1,j_2)k}$.

Finally, the interaction of CC factors' variance component values and IUCC values affected the bias of the estimates, [F(1, 23,976) = 7,491.66, p < .001, = 0.238]. For conditions of unequal variance values, the average bias was lower for IUCC values of 7% (M = 0.022) than for 13% (M = 0.054), resulting in a mean bias difference of 0.032. For conditions of equal variance values, the average bias was lower for IUCC values of 7% (M = 0.047) than for 13% (M = 0.126), resulting in a mean bias difference of 0.079. This interaction is presented in Figure 6. The difference in average bias estimates between the two IUCC values was largest when the variance values were generated to be equal. Overall, as the variance values changed from unequal to equal the mean bias difference between the two IUCC values also increased, from 0.032 to 0.079, a 0.047 difference.

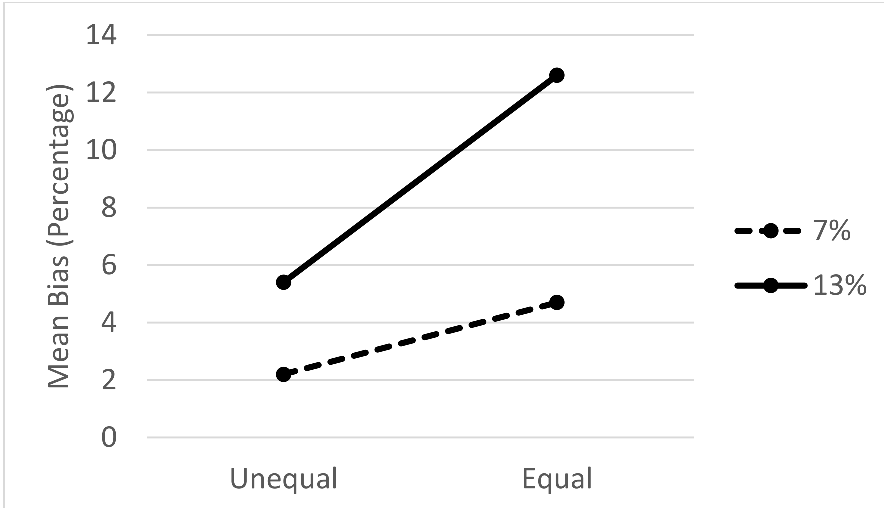

Figure 6. Variance Values and IUCC Values Interaction Effect on the Mean Bias Percentage of $\sigma^2_{i(j_1,j_2)k}$.

Level-2 Variance, $\tau_{0j_1 0k}$

Relative parameter bias for the level-2 variance component, $\tau_{0j_1 0k}$ are presented in Table 12. For *Model 1*, across all conditions, no substantial bias was found and absolute values for the bias remained at less than 0.05. Substantial positive bias was found for all of the 24 conditions in *Model 2*, where the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was not estimated. For *Model 2*, positive parameter bias ranged from 0.152 to 0.443, indicating that the parameter, $\tau_{0j_1 0k}$, was overestimated by 15.2% to 44.3%. Using a practical significance cutoff value of 0.01 for the η^2_p, results for the ANOVA indicated that two main effects (cross-classification structure, and cross-classified factors' variance values) and one two-way interaction affected the severity of the bias of the estimates of $\tau_{0j_1 0k}$.

Table 12

Relative Parameter Bias for the Level-2 Variance Component, τ_{0j_10k}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.016	**0.426**
			13%	0.013	**0.433**
		Unequal	7%	0.014	**0.215**
			13%	0.011	**0.219**
	4 feeder	Equal	7%	0.027	**0.267**
			13%	0.017	**0.255**
		Unequal	7%	0.028	**0.150**
			13%	0.019	**0.139**
Average	2 feeder	Equal	7%	0.005	**0.429**
			13%	0.006	**0.435**
		Unequal	7%	0.004	**0.215**
			13%	0.005	**0.219**
	4 feeder	Equal	7%	0.034	**0.312**
			13%	0.021	**0.293**
		Unequal	7%	0.036	**0.177**
			13%	0.023	**0.161**
Large	2 feeder	Equal	7%	0.016	**0.441**
			13%	0.012	**0.443**
		Unequal	7%	0.014	**0.225**
			13%	0.011	**0.226**
	4 feeder	Equal	7%	0.026	**0.294**
			13%	0.017	**0.280**
		Unequal	7%	0.029	**0.164**
			13%	0.018	**0.152**

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

First, for *Model 2*, the cross-classification structure of the data, 2 feeders or 4 feeders, affected the bias of the estimates of τ_{0j_10k} [F(1, 23,976) = 2,558.40, $p < .001$, η_p^2 = 0.096]. Overall, the average bias was lower for the 4 feeder conditions (M = 0.220) than for the 2 feeder conditions (M = 0.327). On average, across all conditions, estimates of

τ_{0j_10k} were more biased when the data structure less partially cross-classified (and students were only sent to 2 classrooms) than when the data structure became more cross-classified (and students were sent to 4 classrooms).

The cross-classified factors' variance component conditions, all equal or unequal, also affected the bias of the estimates of τ_{0j_10k} [F(1, 23,976) = 6,538.22, p < .001, η_p^2 =0.214]. Overall, the average bias was lower when the level-2 variance component $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was generated to be unequal to the CC factors' variance values, τ_{0j_10k} and τ_{00j_2k}, (M = 0.118) than when all three CC factors' variance values were generated to be equal (M = 0.359). On average, across all other conditions, estimates of τ_{0j_10k} were less biased when the value for the level-2 variance component $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was generated to be less than the value of the two CC factors' variances.

Although, two of the factors in the ANOVA were found to affect the severity of the bias of the τ_{0j_10k} estimates, further analysis of the two-way interactions indicated that the effect of one of the factors depended on the other factor.

For *Model 2*, the interaction of cross-classification structure and CC factors' variance component values affected the bias of the estimate of τ_{0j_10k} [F(1, 23,976) = 435.11, p < .001, η_p^2 = 0.018]. For conditions of 2 feeders, the average bias was lower for unequal values (M = 0.220) than for equal values (M = 0.435), resulting in a mean bias difference of 0.215. For conditions of 4 feeders, the average bias was lower for unequal values (M = 0.157) than for equal values (M = 0.284), resulting in a mean bias difference of 0.127. This interaction is presented in Figure 7. The difference in mean bias between the two variance value conditions was smallest there were 4 feeders. Overall, as the cross-classification became more complete the mean bias difference between the two CC factors' variance component values decreased, from 0.215 to 0.127, a 0.088 difference.

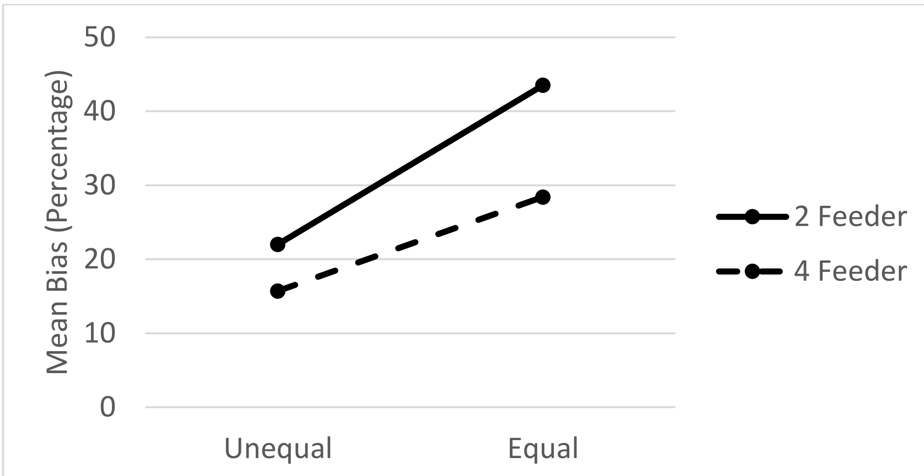

Figure 7. Cross-classification Structure and Variance Values Interaction Effect on the Mean Bias Percentage of τ_{0j_10k}.

Level-2 Variance, τ_{00j_2k}

Relative parameter bias for the level-2 variance component, τ_{00j_2k} are presented in Table 13. For *Model 1*, across all conditions, no substantial bias was found and absolute values for the bias remained at less than .05. Substantial positive bias was found for all of the 24 conditions in *Model 2*, where the variance component, $\tau_{0(j_1 X j_2)k}$ was not estimated. For *Model 2*, positive parameter bias ranged from 0.139 to 0.443, indicating that the parameter, τ_{00j_2k}, was overestimated by 13.9% to 44.3%. Using a practical significance cutoff value of 0.01 for the η_p^2, results for the ANOVA were almost identical to the ANOVA results for the estimation of the other level-2 variance component, $\tau_{0j_1 0k}$. These results indicated that two main effects (cross-classification structure, and cross-classified factors' variance values) and one two-way interaction affected the severity of the bias.

Table 13

Relative Parameter Bias for the Level-2 Variance Component, τ_{00j_2k}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.008	**0.417**
			13%	0.005	**0.425**
		Unequal	7%	0.004	**0.206**
			13%	0.003	**0.211**
	4 feeder	Equal	7%	0.006	**0.255**
			13%	0.006	**0.242**
		Unequal	7%	0.017	**0.139**
			13%	0.008	**0.127**
Average	2 feeder	Equal	7%	0.006	**0.430**
			13%	0.005	**0.434**
		Unequal	7%	0.004	**0.215**
			13%	0.003	**0.217**
	4 feeder	Equal	7%	0.012	**0.290**
			13%	0.003	**0.274**
		Unequal	7%	0.016	**0.157**
			13%	0.006	**0.143**
Large	2 feeder	Equal	7%	0.002	**0.428**
			13%	0.002	**0.433**
		Unequal	7%	0.001	**0.212**
			13%	0.000	**0.216**
	4 feeder	Equal	7%	0.013	**0.281**
			13%	0.004	**0.267**
		Unequal	7%	0.016	**0.152**
			13%	0.006	**0.139**

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

First, for *Model 2*, the cross-classification structure of the data, 2 feeders or 4 feeders, affected the bias of the estimates of τ_{00j_2k} [F(1, 23,976) = 2,969.82, p < .001, η_p^2 = 0.110]. Overall, the average bias was lower for conditions of 4 feeders (M = 0.205) than for 2 feeders (M = 0.320). On average, across all conditions, estimates of τ_{00j_2k} were more biased when the data structure was less cross-classified (and students were only sent to 2 classrooms) than when the data structure became more cross-classified (and students were sent to 4 classrooms).

The cross-classified factors' variance component conditions, all equal or unequal, also affected the bias of the estimates of τ_{00j_2k} [F(1, 23,976) = 6,511.86, p < .001, η_p^2 = 0.214]. Overall, the average bias was lower when the level-2 variance component τ_{00j_2k} was generated to be unequal to the CC factors' variance values, τ_{0j_10k} and τ_{00j_2k}, (M = 0.178) than when all three CC factors' variance values were generated to be equal (M = 0.348). On average, across all other conditions, estimates of τ_{00j_2k} were less biased when the value for the level-2 variance component $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was generated to be less than the value of the two CC factors' variances.

Although, two of the factors in the ANOVA were found to affect the severity of the bias of the τ_{00j_2k} estimates, further analysis of the two-way interactions indicated that the effect of one of the factors depended on the other factor.

The interaction of cross-classification structure and CC factors' variance component values affected the bias of the estimate of τ_{00j_2k} [F(1, 23,976) = 452.870, p <

.001, η_p^2 = 0.019]. For conditions of 2 feeders, the average bias was lower for unequal values (M = 0.213) than for equal values (M = 0.428), resulting in a mean bias difference of 0.215. For conditions of 4 feeders, the average bias was lower for unequal values (M = 0.143) than for equal values (M = 0.268), resulting in a mean bias difference of 0.125. This interaction is presented in Figure 8. The difference in mean bias between the two variance value conditions was smallest when students were sent to 4 classrooms. Overall, as the cross-classification became more complete the mean bias difference between the two CC factors' variance component values decreased, from 0.215 to 0.125, a 0.090 difference.

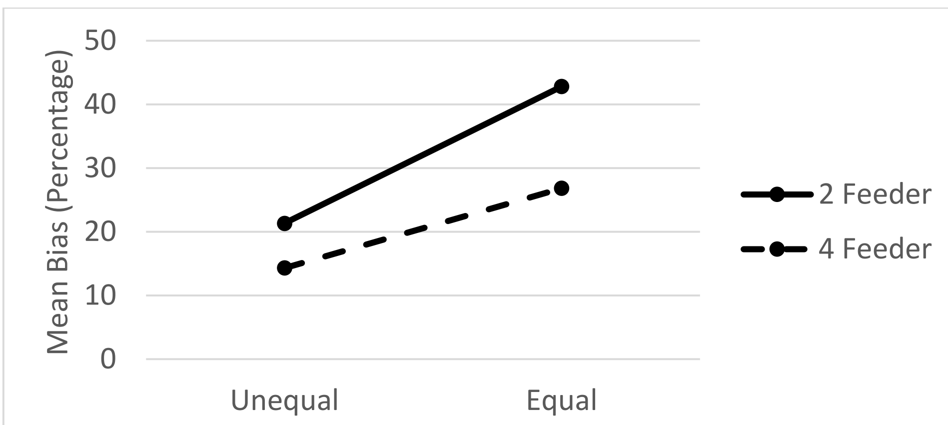

Figure 8. Cross-classification Structure and Variance Values Interaction Effect on the Mean Bias Percentage of τ_{00j2k}.

Level-2 Variance (interaction), $\tau_{0(j_1 \, X \, j_2)k}$

Relative parameter bias for the level-2 variance component associated with the crossed factors' random effects' interaction, $\tau_{0(j_1 \, X \, j_2)k}$ are presented in Table 14. No substantial parameter bias was found for any of the conditions and an ANOVA was not conducted. Only estimates for *Model 1* are included because *Model 2* did not estimate this variance component by design. Across all conditions, relative bias ranged from -0.041 to .015, well below the substantial bias cut-off. Overall, parameter estimates for $\tau_{0(j_1\ X\ j_2)k}$ were not substantially affected by any of the condition specifications.

Table 14

Relative Parameter Bias for the Level-2 Variance Component, $\tau_{0(j_1\ X\ j_2)k}$*, by Condition*

Condition				Estimating Model
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*
Small	2 feeder	Equal	7%	-0.029
			13%	-0.021
		Unequal	7%	-0.041
			13%	-0.029
	4 feeder	Equal	7%	0.006
			13%	0.002
		Unequal	7%	0.013
			13%	0.004
Average	2 feeder	Equal	7%	-0.008
			13%	-0.010
		Unequal	7%	-0.006
			13%	-0.009
	4 feeder	Equal	7%	0.007
			13%	0.002
		Unequal	7%	0.015
			13%	0.006
Large	2 feeder	Equal	7%	-0.013
			13%	-0.012
		Unequal	7%	-0.014
			13%	--0.013
	4 feeder	Equal	7%	0.004
			13%	0.001
		Unequal	7%	0.010
			13%	0.003

Note. The $\tau_{0 j_1 x j_2 k}$ parameter was estimated under *Model 1*. Values with substantial bias are in bold.

Level-3 Variance, τ_{000k}

Relative parameter bias for the level-3 variance component, τ_{000k}, are presented in Table 15. No substantial parameter bias was found for any of the conditions and an ANOVA was not conducted. Relative parameter bias were generally positive for *Model 1* (0.009 to 0.032) and negative for *Model 2* (-0.032 to 0.010, although all estimates were well below the substantial bias cut-off. Overall, parameter estimates for τ_{000k} were not substantially affected by the estimating model type or any of the condition specifications.

Total Variance for All Levels

To provide another depiction of the *Model 2* estimation trends for the variance components, an "aggregated" relative parameter bias was calculated from the *three* random effect variance components generated at level-2: the first cross-classified factor's variance, τ_{0j_10k}, the second CC factors' variance, τ_{00j_2k}, and the variance associated with the interaction of these two crossed factors' random effects, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. That is, a single bias value was calculated to investigate the estimation trends that were occurring at level-2. Relative parameter bias was already calculated for each of these components, separately, but this new value provided a measure of overestimation or underestimation that may have occurred for the entire level. Additionally, a single bias value was calculated for the total variance of level-1, level-2, and level-3. This total variance bias provided a measure of overestimation and underestimation that was occurring for the entire group of variance components at all three levels.

Table 15

Relative Parameter bias for the Level-3 Variance Component, τ_{000k} , by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.015	-0.016
			13%	0.030	-0.030
		Unequal	7%	0.014	-0.001
			13%	0.030	0.001
	4 feeder	Equal	7%	0.013	-0.007
			13%	0.028	-0.007
		Unequal	7%	0.012	0.003
			13%	0.028	0.010
Average	2 feeder	Equal	7%	0.017	-0.015
			13%	0.032	-0.029
		Unequal	7%	0.017	0.001
			13%	0.032	0.002
	4 feeder	Equal	7%	0.012	-0.007
			13%	0.027	-0.006
		Unequal	7%	0.011	0.002
			13%	0.026	0.009
Large	2 feeder	Equal	7%	0.013	-0.019
			13%	0.029	-0.032
		Unequal	7%	0.013	-0.003
			13%	0.029	-0.002
	4 feeder	Equal	7%	0.009	-0.010
			13%	0.026	-0.010
		Unequal	7%	0.009	-0.001
			13%	0.024	0.006

Note. The $\tau_{0\,j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 16 presents the relative parameter bias calculated for the estimation of all the variance components, broken down by each level and including the total level-2 random effects' variance and the total variance for all levels. For *Model 2*, the misspecified model which involved the assumption that $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was zero when it was not zero, bias for the

Table 16

Relative Parameter Bias for All Variance by Condition

Condition				Variance Estimates					
Size	Partial CC structure	CC factors' variance values	IUCC	L-1	L-2 Kinder	L-2 First	L-2 **Total**	L-3	Overall Total
Small	2 feeder	Equal	7%	0.033	**0.426**	**0.417**	**-0.053**	-0.016	0.005
			13%	**0.086**	**0.433**	**0.425**	**-0.048**	-0.030	0.011
		Unequal	7%	0.016	**0.215**	**0.206**	-0.033	-0.001	0.004
			13%	0.037	**0.219**	**0.211**	-0.029	0.001	0.008
	4 feeder	Equal	7%	0.068	**0.267**	**0.255**	**-0.160**	-0.007	0.005
			13%	0.182	**0.255**	**0.242**	**-0.168**	-0.007	0.008
		Unequal	7%	0.032	**0.150**	**0.139**	**-0.085**	0.003	0.005
			13%	**0.079**	**0.139**	**0.127**	**-0.094**	0.010	0.009
Average	2 feeder	Equal	7%	0.032	**0.429**	**0.430**	-0.048	-0.015	0.006
			13%	**0.085**	**0.435**	**0.434**	-0.044	-0.029	0.012
		Unequal	7%	0.015	**0.215**	**0.215**	-0.029	0.001	0.005
			13%	0.037	**0.219**	**0.217**	-0.026	0.002	0.009
	4 feeder	Equal	7%	**0.058**	**0.312**	**0.290**	**-0.133**	-0.007	0.005
			13%	**0.155**	**0.293**	**0.274**	**-0.145**	-0.006	0.006
		Unequal	7%	0.027	**0.177**	**0.157**	**-0.067**	0.002	0.005
			13%	**0.067**	**0.161**	**0.143**	**-0.079**	0.009	0.008
Large	2 feeder	Equal	7%	0.031	**0.441**	**0.428**	-0.045	-0.019	0.005
			13%	**0.083**	**0.443**	**0.433**	-0.042	-0.032	0.011
		Unequal	7%	0.015	**0.225**	**0.212**	-0.026	-0.003	0.004
			13%	0.036	**0.226**	**0.216**	-0.024	-0.002	0.009
	4 feeder	Equal	7%	**0.061**	**0.294**	**0.281**	**-0.142**	-0.010	0.003
			13%	**0.163**	**0.280**	**0.267**	**-0.151**	-0.010	0.006
		Unequal	7%	0.028	**0.164**	**0.152**	**-0.075**	-0.001	0.004
			13%	**0.070**	**0.152**	**0.139**	**-0.084**	0.006	0.005

Note. The $\tau_{0 j_1 x j_2 k}$ variance was not estimated under *Model 2*. Values with substantial bias are in bold. Size = classroom sample size; L-1 = Level-1; L-2 = Level-2; L-3 = Level-3; All = Total mean bias for all levels.

estimations of the total level-2 variance was negative for all conditions. Across all conditions, the total level-2 variance was underestimated when *Model 2* was used for estimation. In 14 of the conditions, the negative bias was considered substantial. These bias values ranged from -0.024 to -0.151, indicating that this variance was underestimated by 2.4% to 15.1%. Results for the ANOVA indicated that two main effects (cross-classification structure, and cross-classified factors' variance values) and one two-way interaction affected the severity of the bias of the total level-2 variance.

First, the cross-classification structure of the data, 2 feeders or 4 feeders, affected the bias of the estimates of τ_{00j_2k} [F(1, 23,976) = 5,520.50, p < .001, η_p^2 = 0.187]. Overall, the average bias was larger for conditions with 4 feeders (M = -0.115) than for 2 feeders (M = -0.037). On average, across all conditions, estimates of the total level-2 variance were less biased when the data structure was less cross-classified (and students were only sent to 2 classrooms) than when the data structure was more cross-classified (and students were sent to 4 classrooms).

The cross-classified factors' variance component conditions, all equal or unequal, also affected the bias of the estimates of τ_{00j_2k} [F(1, 23,976) = 1,824.86, p < .001, η_p^2 = 0.071]. Overall, the average bias was smaller when the level-2 variance component τ_{00j_2k} was unequal to the cross-classified factors' variance values, τ_{0j_10k} and τ_{00j_2k}, (M = -0.054) than when all three CC factors' variance values were generated to be equal (M = -0.099). On average, across all other conditions, estimates of the total level-2 variance were less biased when the value for the level-2 variance component was generated to be less than the value of the two cross-classified factors' variances.

The interaction of cross-classification structure and the cross-classified factors' variance component values affected the bias of the total level-2 variance estimates [F(1, 23,976) = 570.55, p < .001, η_p^2 =0.023]. For conditions of 2 feeders, the average bias was lower for unequal values (M = -0.027) than for equal values (M = -0.047), resulting in a mean bias difference of 0.020. For conditions of 4 feeders, the average bias was lower for unequal values (M = -0.080) than for equal values (M = -0.150), resulting in a mean bias difference of 0.070. The difference in mean bias between the two variance value conditions was largest when data become more cross-classified. This interaction is presented in Figure 9. Overall, as data structure became more cross-classified the mean bias difference between the two cross-classified factors' variance component values increased, from 0.020 to 0.070, a 0.050 difference.

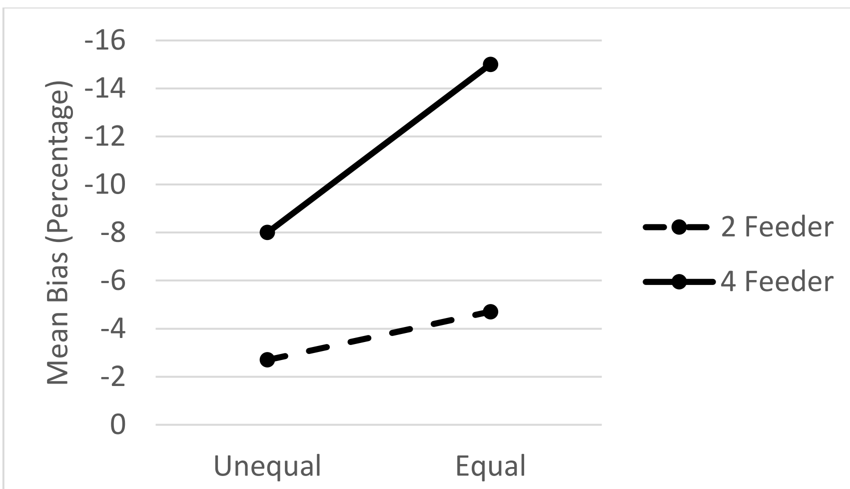

Figure 9. Cross-classification Structure and Variance Values Interaction Effect on the Mean Bias Percentage of the Total Level-2 Variance.

While the total level-2 variance estimates were negatively biased, the individual level-2 variances, τ_{0j_10k} and τ_{00j_2k}, were consistently overestimated. The level-1 component, as well, showed positive bias, while the level-3 variance had some underestimation, but it was not substantial.

For the total variance of all levels, bias for the variance estimates were at 0.01 or below across all conditions. No substantial bias was occurring for the total variance in any condition and the total variance estimations for *Model 2* were less than 1% overestimated across all conditions.

RELATIVE STANDARD ERROR BIAS

Fixed Effects

Relative standard error (SE) bias was calculated for four fixed effect estimates: the coefficient of the student level predictor, γ_{1000}, the coefficients for the two level-2 predictors associated with the crossed factors, γ_{0100} and γ_{0010} , and the intercept at level-1, γ_{0000}. These bias were estimated for *Model 1*, the correct model that included the estimation of $\tau_{0(j_1,j_2)k}$ (which had been generated to be non-zero) and *Model 2*, the misspecified model, which involved an assumption that $\tau_{0(j_1,j_2)k}$ was zero and is also most commonly used by applied researchers estimating a CCREM. *SE* estimates were considered to be substantially biased if the absolute value of the relative *SE* bias was greater than or equal to 0.10 (Hoogland & Boomsma, 1998). ANOVAs were conducted when substantial *SE* bias was found for multiple conditions.

Intercept, γ_{0000}

Relative *SE* bias values for estimates of the intercept, γ_{0000}, are presented in Table 17. This table summarizes bias for both the *Model 1* and *Model 2* estimates in each of the 24 conditions. No substantial *SE* bias was found for any of the conditions, therefore an ANOVA was not conducted. Within the same condition, *SE* estimates were similarly biased between the two estimating models. Values ranged from -0.006 to 0.027, well below the acceptable cutoff of .10 (Hoogland & Boomsma, 1998). Overall, these *SE* bias were not substantially affected by the estimating model, class size, cross-classification structure, the cross-classified (CC) factors' variance values, or the IUCC values.

Student Level Predictor Coefficient, γ_{1000}

Relative *SE* bias for the coefficient of the level-1 predictor, γ_{1000}, are presented in Table 18. No substantial relative *SE* bias was found for any of the conditions, and again, a follow up ANOVA was not conducted. Within the same conditions, *SE* estimates were more biased for *Model 2* than *Model 1* when classroom sizes were small. In general, bias decreased as classroom size conditions went from small to large. Ranges for these bias estimates were similar for *Model 1* (-0.033 to 0.053) and *Model 2* (-0.037 to 0.056), both much lower than the cutoff for acceptable bias. Overall, *SE* estimates associated with the coefficient γ_{1000} were not substantially affected by any of the condition specifications or estimating models.

Table 17

Relative SE Bias for the Intercept, γ_{0000}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.018	0.020
			13%	0.019	0.020
		Unequal	7%	0.017	0.018
			13%	0.019	0.020
	4 feeder	Equal	7%	-0.002	-0.001
			13%	0.000	0.000
		Unequal	7%	-0.003	-0.003
			13%	0.001	0.001
Average	2 feeder	Equal	7%	0.016	0.015
			13%	0.019	0.018
		Unequal	7%	0.015	0.015
			13%	0.019	0.019
	4 feeder	Equal	7%	0.002	-0.001
			13%	-0.002	-0.003
		Unequal	7%	0.002	0.000
			13%	0.000	-0.002
Large	2 feeder	Equal	7%	0.025	0.024
			13%	0.022	0.022
		Unequal	7%	0.027	0.027
			13%	0.025	0.025
	4 feeder	Equal	7%	-0.002	-0.004
			13%	-0.004	-0.006
		Unequal	7%	0.000	-0.002
			13%	-0.002	-0.003

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 18

Relative SE Bias for the X coefficient, γ_{1000}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	-0.032	-0.037
			13%	-0.028	-0.041
		Unequal	7%	-0.033	-0.035
			13%	-0.030	-0.037
	4 feeder	Equal	7%	0.051	0.056
			13%	0.046	0.052
		Unequal	7%	0.053	0.056
			13%	0.050	0.055
Average	2 feeder	Equal	7%	0.040	0.037
			13%	0.043	0.038
		Unequal	7%	0.038	0.037
			13%	0.042	0.039
	4 feeder	Equal	7%	0.004	0.001
			13%	0.007	-0.002
		Unequal	7%	0.004	0.002
			13%	0.005	0.001
Large	2 feeder	Equal	7%	0.011	0.009
			13%	0.011	0.009
		Unequal	7%	0.011	0.010
			13%	0.011	0.010
	4 feeder	Equal	7%	0.003	0.006
			13%	0.002	0.008
		Unequal	7%	0.004	0.006
			13%	0.004	0.007

Note. The $\tau_{0 j_1 x j_2 k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

***W Coefficient,* γ_{0100}**

The relative *SE* bias for the coefficients of the level-2 predictor, γ_{0100}, are presented in Table 19. This predictor is associated with the first cross-classified factor, here, F1. There was no substantial *SE* bias found for any of the conditions and an ANOVA was not conducted. Within the same conditions, *SE* estimates were similarly biased between *Model 1* and *Model 2*. Across all conditions, relative *SE* bias ranged from -0.001 to 0.056, well below the substantial bias cut-off. Overall, *SE* estimates associated with this coefficient were not substantially affected by the estimating model or any of the conditions.

Table 19

Relative SE Bias for the W coefficient, γ_{0100} *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.056	0.055
			13%	0.055	0.055
		Unequal	7%	0.049	0.048
			13%	0.049	0.049
	4 feeder	Equal	7%	0.007	0.010
			13%	0.008	0.010
		Unequal	7%	0.012	0.014
			13%	0.012	0.014
Average	2 feeder	Equal	7%	0.024	0.023
			13%	0.030	0.029
		Unequal	7%	0.019	0.017
			13%	0.024	0.023
	4 feeder	Equal	7%	0.001	0.006
			13%	-0.001	0.004
		Unequal	7%	0.005	0.007
			13%	0.003	0.007
Large	2 feeder	Equal	7%	0.042	0.040
			13%	0.044	0.043
		Unequal	7%	0.037	0.036
			13%	0.039	0.038
	4 feeder	Equal	7%	0.007	0.004
			13%	0.006	0.003
		Unequal	7%	0.012	0.011
			13%	0.011	0.010

Note. The $\tau_{0\,j_1 x j_2 k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Z Coefficient, γ_{0010}

The relative *SE* bias for the coefficients of the other level-2 predictor, γ_{0010}, are presented in Table 20. This predictor is associated with the second cross-classified factor, here, F2. There was no substantial *SE* bias was found for any conditions and an ANOVA was not necessary. Within the same conditions, *SE* estimates were similarly biased between *Model 1* and *Model 2*. Across all conditions, relative bias ranged from -0.015 to 0.031, well below the substantial bias cut-off. Overall, *SE* estimates associated with the coefficient γ_{0010}, were not substantially affected by estimating model or any of the condition specifications.

Random Effects

Relative *SE* bias was calculated for five random effect variance component estimates: the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$, the two cross-classified factors' variance values, τ_{0j_10k} and τ_{00j_2k}, the level-2 variance component associated with the interaction of the crossed factors' random effects, $\tau_{0(j_1 X j_2)k}$, and the level-3 variance component, τ_{000k}. As above, *SE* estimates were considered to be substantially biased if the absolute value of the relative *SE* bias was greater than or equal to 0.10 (Hoogland & Boomsma, 1998). ANOVAs were conducted when substantial *SE* bias was found for multiple conditions.

Table 20

Relative SE Bias for the Z coefficient, γ_{0010} *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	-0.015	-0.011
			13%	-0.012	-0.007
		Unequal	7%	-0.015	-0.013
			13%	-0.010	-0.008
	4 feeder	Equal	7%	0.007	0.007
			13%	0.002	0.004
		Unequal	7%	0.007	0.007
			13%	0.003	0.004
Average	2 feeder	Equal	7%	0.028	0.030
			13%	0.017	0.019
		Unequal	7%	0.030	0.031
			13%	0.022	0.023
	4 feeder	Equal	7%	-0.005	-0.003
			13%	-0.010	-0.007
		Unequal	7%	-0.005	-0.005
			13%	-0.008	-0.006
Large	2 feeder	Equal	7%	0.016	0.020
			13%	0.005	0.009
		Unequal	7%	0.024	0.027
			13%	0.013	0.015
	4 feeder	Equal	7%	-0.008	-0.007
			13%	-0.012	-0.012
		Unequal	7%	-0.006	-0.006
			13%	-0.009	-0.011

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Level-1 Variance, $\sigma^2_{i(j_1,j_2)k}$

The relative *SE* bias for estimates of the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$, are presented in Table 21. For *Model 1*, there was no substantial bias found, while *Model 2* had one condition with a mean bias value of -0.104, just slightly larger than the acceptable cutoff. This was the only condition with substantial bias, so an ANOVA was not conducted. For *Model 1*, *SE* bias were in the positive range (0.009 to 0.034), while these bias were both positive and negative for *Model 2* estimates (-0.104 to 0.030).

Table 21

Relative SE Bias for the Level-1 Variance Component, $\sigma^2_{i(j_1,j_2)k}$ *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	0.023	0.020
			13%	0.023	-0.052
		Unequal	7%	0.023	0.030
			13%	0.023	0.015
	4 feeder	Equal	7%	0.034	0.006
			13%	0.032	-0.053
		Unequal	7%	0.035	0.017
			13%	0.033	0.001
Average	2 feeder	Equal	7%	0.012	0.011
			13%	0.012	-0.076
		Unequal	7%	0.012	0.024
			13%	0.012	0.006
	4 feeder	Equal	7%	0.016	-0.004
			13%	0.019	-0.044
		Unequal	7%	0.014	-0.001
			13%	0.018	-0.005
Large	2 feeder	Equal	7%	0.025	0.002
			13%	0.025	**-0.104**
		Unequal	7%	0.025	0.022
			13%	0.025	-0.006
	4 feeder	Equal	7%	0.010	0.007
			13%	0.011	-0.054
		Unequal	7%	0.009	0.014
			13%	0.010	0.003

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Level-2 Variance, $\tau_{0j_1 0k}$

The relative *SE* bias for the level-2 variance component, $\tau_{0j_1 0k}$, associated with the first crossed factor, F1, are presented in Table 22. Across all conditions, no substantial bias found and an ANOVA was not conducted. *SE* estimates were similarly biased between the two estimating models, with a bias range of (-0.028 to 0.049) for *Model 1* and a range of -0.034 to 0.039 for *Model 2*. Overall, none of the conditions substantially affected the bias.

Level-2 Variance, τ_{00j_2k}

The relative *SE* bias for the level-2 variance component, τ_{00j_2k}, which is associated with the second crossed factor, F2 , are presented in Table 23. Across all conditions, there was no substantial bias found and an ANOVA was not conducted. *SE* estimates were similarly biased between the two estimating models, with a bias range of -0.022 to 0.053 for *Model 1* and a range of -0.020 to 0.048 for *Model 2*. Overall, none of the conditions substantially affected the bias.

Level-2 Variance (interaction), $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$

The relative *SE* bias for the level-2 variance component associated with the crossed factors' random effects' interaction, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, are presented in Table 24. Only bias for the estimates from *Model 1* are included because this variance component was not estimated under *Model 2* by design. Across all conditions, there was no substantial bias found and an ANOVA was not conducted. *SE* estimates bias ranged from -0.027 to 0.026, well below the acceptable cutoff of 0.10.

Table 22

Relative SE Bias for the Level-2 Variance Component, τ_{0j_10k}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	-0.014	0.002
			13%	-0.014	-0.002
		Unequal	7%	-0.014	0.005
			13%	-0.014	0.001
	4 feeder	Equal	7%	0.047	0.035
			13%	0.049	0.039
		Unequal	7%	0.037	0.027
			13%	0.039	0.032
Average	2 feeder	Equal	7%	-0.008	0.007
			13%	-0.005	0.005
		Unequal	7%	-0.016	-0.003
			13%	-0.010	-0.002
	4 feeder	Equal	7%	-0.029	-0.034
			13%	-0.015	-0.019
		Unequal	7%	-0.028	-0.032
			13%	-0.017	-0.020
Large	2 feeder	Equal	7%	0.016	0.024
			13%	0.004	0.011
		Unequal	7%	0.011	0.018
			13%	0.002	0.006
	4 feeder	Equal	7%	0.027	0.029
			13%	0.023	0.025
		Unequal	7%	0.028	0.028
			13%	0.023	0.024

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 23

Relative SE Bias for the Level-2 Variance Component, $\tau_{00 j_2 k}$ *, by Condition*

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	*Model 2*
Small	2 feeder	Equal	7%	-0.001	-0.006
			13%	0.001	-0.004
		Unequal	7%	0.001	-0.008
			13%	-0.002	-0.007
	4 feeder	Equal	7%	0.014	0.005
			13%	-0.006	-0.012
		Unequal	7%	0.020	0.010
			13%	0.002	-0.004
Average	2 feeder	Equal	7%	-0.010	-0.016
			13%	-0.004	-0.006
		Unequal	7%	-0.008	-0.013
			13%	-0.004	-0.005
	4 feeder	Equal	7%	0.012	0.006
			13%	-0.009	-0.020
		Unequal	7%	0.020	0.018
			13%	-0.001	-0.006
Large	2 feeder	Equal	7%	0.049	0.045
			13%	0.038	0.034
		Unequal	7%	0.053	0.048
			13%	0.042	0.037
	4 feeder	Equal	7%	-0.010	0.000
			13%	-0.022	-0.019
		Unequal	7%	-0.008	-0.001
			13%	-0.018	-0.016

Note. The $\tau_{0 j_1 x j_2 k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Table 24

Relative SE Bias for the Level-2 Variance Component, $\tau_{0(j_1\ x\ j_2)k}$*, by Condition*

Condition				Estimating Model
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*
Small	2 feeder	Equal	7%	-0.019
			13%	-0.021
		Unequal	7%	-0.012
			13%	-0.018
	4 feeder	Equal	7%	0.012
			13%	-0.004
		Unequal	7%	0.026
			13%	0.007
Average	2 feeder	Equal	7%	-0.027
			13%	-0.025
		Unequal	7%	-0.026
			13%	-0.026
	4 feeder	Equal	7%	0.018
			13%	0.026
		Unequal	7%	0.015
			13%	0.022
Large	2 feeder	Equal	7%	-0.019
			13%	-0.021
		Unequal	7%	-0.016
			13%	-0.019
	4 feeder	Equal	7%	0.009
			13%	0.009
		Unequal	7%	0.009
			13%	0.011

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Level-3 Variance, τ_{000k}

The relative *SE* bias for the level-3 variance component, τ_{000k}, are presented in Table 25. Overall, no substantial relative *SE* bias was found for any of the conditions, and an ANOVA was not necessary. Across all conditions, *SE* estimates were similarly biased between the two estimating models.

Table 25

Relative SE Bias for the Level-3 Variance Component, τ_{000k}, by Condition

Condition				Estimating Model	
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	*Model 1*	Model 2
Small	2 feeder	Equal	7%	-0.022	-0.023
			13%	-0.020	-0.020
		Unequal	7%	-0.024	-0.024
			13%	-0.023	-0.023
	4 feeder	Equal	7%	-0.010	-0.010
			13%	-0.008	-0.007
		Unequal	7%	-0.012	-0.013
			13%	-0.010	-0.010
Average	2 feeder	Equal	7%	-0.010	-0.011
			13%	-0.010	-0.010
		Unequal	7%	-0.012	-0.012
			13%	-0.011	-0.012
	4 feeder	Equal	7%	-0.008	-0.002
			13%	-0.009	0.000
		Unequal	7%	-0.008	-0.006
			13%	-0.009	-0.004
Large	2 feeder	Equal	7%	-0.004	-0.006
			13%	-0.003	-0.005
		Unequal	7%	-0.008	-0.009
			13%	-0.007	-0.008
	4 feeder	Equal	7%	-0.020	-0.014
			13%	-0.016	-0.007
		Unequal	7%	-0.022	-0.018
			13%	-0.018	-0.013

Note. The $\tau_{0j_1xj_2k}$ parameter was estimated under *Model 1* and it was not estimated under *Model 2*. Values with substantial bias are in bold.

Information criteria

Table 26 presents the information criteria results for the AIC, the AICC, and the BIC information criteria. These are all the default information criteria for SAS PROC MIXED. The table shows the percentage of times out of 1,000 replications that the fit of *Model 1*, the model that included $\tau_{0(j_1 X j_2)k}$ in its estimation, would have been selected as a better fit than *Model 2*. For all the indices, a low value indicates a better fitting model, so here, the information criteria values were directly compared for both Models in order to select the lowest value, which indicated the better fitting model. The AIC and the AICC consistently selected *Model 1* as the better fitting model, with the lowest percentage for a combination of conditions being 93.9%. The BIC was not as consistent, and percentages ranged from 10.20% to 100% of correct model selections.

While an ANOVA was not conducted, the percent of times the BIC selected *Model 2* as the correct model was lowest for the IUCC conditions of 7%. Further, conditions with unequal variances and 7% IUCC had the lowest of the proportions correct. When the IUCC was 13%, the lowest percentage was 96.10%, while when the IUCC was 7%, the lowest percentage was 10.20%. Overall, IUCC and unequal level-2 variances had the most impact on the correct model selection using the BIC.

Table 26

Percent of Correct Model Identification by Condition

Condition				Information criteria			
Classroom sample size	Partial cross-classification structure	CC factors' variance values	IUCC	**Log-Likelihood**	**AIC**	**AICC**	**BIC**
Small	2 feeder	Equal	7%	100.00	100.00	100.00	89.00
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	99.50	94.00	93.90	10.30
			13%	100.00	100.00	100.00	96.10
	4 feeder	Equal	7%	100.00	100.00	100.00	96.40
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	99.90	97.20	97.30	10.20
			13%	100.00	100.00	100.00	99.50
Average	2 feeder	Equal	7%	100.00	100.00	100.00	98.50
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	99.90	98.80	98.80	31.20
			13%	100.00	100.00	100.00	99.70
	4 feeder	Equal	7%	100.00	100.00	100.00	99.80
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	100.00	99.30	99.30	24.60
			13%	100.00	100.00	100.00	100.00
Large	2 feeder	Equal	7%	100.00	100.00	100.00	99.90
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	100.00	99.70	99.70	53.10
			13%	100.00	100.00	100.00	100.00
	4 feeder	Equal	7%	100.00	100.00	100.00	100.00
			13%	100.00	100.00	100.00	100.00
		Unequal	7%	100.00	100.00	100.00	56.70
			13%	100.00	100.00	100.00	100.00

Chapter 5: Discussion

SUMMARY

Two studies investigated estimation of the crossed factors' random effects' interaction, $\tau_{0(j_1\ X\ j_2)k}$ in three level CCREM models. For both of these studies, results for two different estimating models were compared: one model included the variance component, $\tau_{0(j_1\ X\ j_2)k}$, and the second model assumed the value of this variance component was zero. The second model is the model most commonly assumed by researchers utilizing the CCREM to estimate crossed factors' effects. The first study involved use of a real world data set, the STAR data, and compared parameter estimates for both estimating models. The results for this study served as a guide to provide generating parameter values for the Monte Carlo simulation study that followed. The second study, a simulation study, also investigated the same two estimating models, but the simulation allowed for manipulation of different conditions, as well as, allowing for measures of parameter and standard error bias. Findings for the real world data analysis (the first study) were already discussed in the Methods section, as their results were necessary to conduct the simulation study. This chapter will focus on the simulation study, its results, limitations, potential for future research, and impact on education.

Fixed Effects

Overall, no substantial bias was found for the fixed effect estimates' for *Model 1* (which estimated $\tau_{0(j_1\ X\ j_2)k}$) or *Model 2* (which assumed $\tau_{0(j_1\ X\ j_2)k}$ was 0). Across all conditions, neither estimating model produced substantial bias in estimates for the fixed effects' coefficients. Overall, when conditions are similar to those found in this study, using either of the two models to estimate parameters in a given dataset should not result in substantially biased parameter estimates for the fixed effects. That is, if researchers continue to model the variance component, $\tau_{0(j_1 X j_2)k}$, as zero, even when its true value is not zero, and the dataset has similar characteristics to those simulated here, parameter estimates of the fixed effects should not have substantial bias.

Additionally, for the estimating models used in the simulation, no substantial bias was found in the associated standard error (SE) estimates of the fixed effects. That is, relative SE bias estimates for *Model 1* and *Model 2* were not found to exceed an absolute value of 10% in any of the conditions simulated. These findings are consistent with those of Shi et al. (2010) who also found no substantial parameter or standard error bias for the fixed effect estimates (and their associated SEs), when $\tau_{0(j_1 X j_2)k}$ was generated to be non-zero, but was **not** included in the estimating model (similar to *Model 2* in the current simulation study).

Although they were not specifically investigating the variance component, $\tau_{0(j_1 X j_2)k}$, Luo and Kwok (2009) also had similar findings for the fixed effects' estimates when using a misspecified model to estimate cross-classified data structures. They found that when an estimating model did not model a cross-classified factor *and* that factor was used to generate the data, then the parameter estimates and associated standard errors of the fixed effects were not affected. These findings were consistent across several of their simulation studies, regardless of the cross-classification structure of their data (partial to full complete cross-classification). While both of the estimating models in the current simulation study did appropriately model both cross-classified factors that were generated, *Model 2* was misspecified because $\tau_{0(j_1 \; X \; j_2)\mathrm{k}}$ (associated with both crossed factors) was not zero and was not estimated, as it was assumed to be zero. Similar to Luo and Kwok, parameter and SE estimates for the fixed effects' were not substantially biased for the misspecified model. This pattern did not change as the data became more cross-classified.

Conditions in the simulation included three classroom sample sizes for the first cross-classified factor, for which students were selected from and assigned to a specific classroom for the second cross-classified factor. Relative parameter bias of the fixed effect coefficients and their associated SEs, were not substantial, regardless of the classroom sample size: small, average, or large. Two values of IUCC were also included in this study: 7% and 13%. The IUCC's were associated with the level-2 factors, which were simulated to be kindergarten and first grade classrooms. An IUCC increase from 7% to 13% meant that the variance between classrooms increased from 7% to 13%. An increase in the IUCC value did not have a substantial impact on the bias of the fixed effect coefficients and their associated SEs, regardless of whether $\tau_{0(j_1 \; X \; j_2)\mathrm{k}}$ was included in the estimating model. Additionally, parameter bias and SE bias for the fixed effect estimates were not substantially impacted when $\tau_{0(j_1 \; X \; j_2)\mathrm{k}}$ was manipulated to either equal or less than the two crossed factor variance values, $\tau_{0 j_1 0k}$ and $\tau_{00 j_2 k}$.

Overall, findings for the relative parameter bias for the fixed effect estimates, and bias for their associated SEs were similar to previous research (Luo & Kwok, 2009; Shi et al., 2010) investigating similar cross-classified data structures. No substantial SE or parameter bias was found for the fixed effect estimates when the estimating model was misspecified (similar to *Model 2* here), and a variance component associated with a cross-classified factor was not included in the estimation.

Random Effect Variance Component Estimates

Unlike the fixed effect estimates described above, substantial bias was found for some of the random effect variance component estimates and their associated standard errors. This bias was not found for *Model 1*, which correctly estimated $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, but the variance estimates were found to be biased when *Model 2* was estimated, which did not include the estimation of the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. Discussions for each variance component is done in separate sections in order to maintain clarity.

Level-1 Variance, $\sigma^2_{i(j_1,j_2)k}$

Substantial bias was found for *Model 2*'s estimation of the level-1 variance component. Estimates for *Model 1*, which correctly modeled $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ at level-2, resulted in an insubstantial amount of bias for the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$. Estimates for *Model 2*, on the other hand, resulted in substantial positive bias for half of the simulated conditions. That is, when the level-2 variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, was generated to be non-zero, but was not estimated in the estimating model (*Model 2*), the variance at level-1 was *overestimated*. This is similar to what was found in the study by Luo and Kwok (2009), where misspecification of a three-level CCREM model with cross-classification at the intermediate level did result in slight positive bias for the level-1 variance estimations. For the present study, overestimation of the level-1 variance may have been impacted by several of the simulated conditions.

Estimates of the level-1 variance (from *Model 2*) were more positively biased when the data structure was more cross-classified. In this study, the cross-classification of the data structure was determined by the number of classrooms that were considered “feeders”. When students were “fed” into two classrooms, the cross-classification was considered less cross-classified. When students were “fed” into four classrooms, the cross-classification was considered more cross-classified. As the data structure became more cross-classified, overestimation of this variance component was larger. This corresponds to bias patterns in Luo and Kwok (2009) where misspecification of a variance component at level-2 may result in positive parameter bias for the level-1 variance component, which increases in magnitude as the data structure becomes more cross-classified. For the present simulation, the partial eta squared value (.484) indicated a significant practical effect of the cross-classification structure on the bias of the level-1 variance estimates. When comparing the mean bias for both of the cross-classification conditions, however, the difference was .041, or 4.1%. This is a relatively small percentage difference, despite the partial eta squared value.

The impact of the cross-classification structure on the bias of the level-1 variance component estimates depended on 3 other conditions in the model. The positive bias observed in the 4 feeder condition varied as the IUCC values varied, such that more bias was observed for the 4 feeders than the 2 feeder condition when IUCC values were larger (13%). That is, when the proportion of variability attributed to the classroom factor at level-2 is larger, the level-1 variance is more likely to be overestimated, an overestimation that increases as data structure at level-2 becomes more cross-classified. The partial eta squared value (η_p^2 = 0.162) indicated a significant practical effect of this interaction on the bias of the level-1 variance estimates. Overall, as the IUCC changed from 7% to 13%, the mean bias difference between the cross-classification structures increased by 0.036, or 3.6%, a relatively small percentage difference.

The positive bias observed in the cross-classified structure conditions varied by classroom sample size. When the classroom sample size was the smallest, the level-1 variance was more overestimated when students were sent to 4 classrooms. This overestimation pattern occurred for all classroom sizes, but differences between the 2 and 4 feeder conditions were the largest for the smallest classroom size. The partial eta squared value (η_p^2 = 0.016) indicated a significant practical effect of this interaction on the bias of the level-1 variance estimates. Further examination, however, showed that the mean bias difference between the 2 feeder and 4 feeder conditions was only 0.012, or 1.2% larger for the smaller classroom size than for the average classroom size.

Overestimation of the level-1 variance attributed to the cross-classification structure also depended on the values of the level-2 variance components. Estimates of the level-1 variance (from *Model 2*) were also *more* positively biased when the level-2 variance components were all set to be equal in value ($\tau_{0(j_1\ X\ j_2)\mathrm{k}} = \tau_{0\,j_1 0k} = \tau_{00\,j_2 k}$). Conversely, when the level -2 variances were unequal ($\tau_{0(j_1\ X\ j_2)\mathrm{k}} \neq \tau_{0\,j_1 0k}$ nor $\tau_{00\,j_2 k}$) and, therefore, the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, was a smaller value than the other level-2 variance components, there was less overestimation of the level-1 variance component. This overestimation depended on the cross-classification structure at level-2. More overestimation occurred for the equal values condition than unequal conditions when the data was more cross-classified (4 feeders). The partial eta squared value (η_p^2 = 0.13) indicated a significant practical effect of this interaction on the bias of the level-1 variance estimates. Overall, however, as the conditions changed from 2 feeders to 4 feeders, the mean bias difference between the equal/unequal values increased by 0.032, or 3.2%. This is a relatively small percentage increase, although partial eta squared was a larger value.

Overestimation of the level-1 variance attributed to the value of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ also depended on the IUCC values for the level-2 cross-classified factors. For conditions where all the level-2 variances are equal, overestimation of the level-1 variance increased as the values of the IUCC increased. This means that when the proportion of variability attributed to the classroom factor at *level-2* (IUCC) was larger, the level-1 variance is more likely to be overestimated, an overestimation that increases when the level-2 variances are equal (and subsequently, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is a larger value). Shi et al (2006) did manipulate the value of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, and found that the value of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ had no substantial impact on bias for the level-1 variance. While Shi et al. (2010) looked at two different values of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, these two values were equal to the values of the other level-2 variance components. That is, their study had all three level-2 variance values equal to each other in all conditions. The present study introduced a condition where these variances were not equal, and therefore the value of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was smaller than the other two level-2 variances. As observed here, this condition of equal/not equal did impact the level-1 variance estimation bias, and may be able to explain why this same impact was not found for Shi et al.

Although some negative bias was observed for the SE associated with the level-1 variance component, only one condition found substantial negative bias. This only occurred for *Model 2* estimates of the condition with a large IUCC value, equal variance values, and 2 feeders, for students in large classroom sizes. Although similar conditions were not manipulated, the underestimation for this condition corresponds to findings in Meyers and Beretvas (2006) where misspecification of a variance component at level-2 may result in negative bias for the level-1 variance component's associated standard error. Further investigation is needed to determine under what conditions the underestimation of the standard errors are impacted when misspecification is related to the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

Level-2 Variances, τ_{0j_10k} ***and*** τ_{00j_2k}

For the simulated data in the present study, level-2 is the level where the cross-classification was simulated to occur. The level-2 variance components, τ_{0j_10k} and τ_{00j_2k}, are associated with the two cross-classified factors. Patterns of bias were almost identical for the two variances and are therefore discussed together in this section.

Substantial bias was found for *Model 2*'s estimation of the two level-2 variance components, τ_{0j_10k} and τ_{00j_2k}. Estimates for *Model 1,* which correctly modeled $\tau_{0(j_1 X j_2)\text{k}}$ at level-2, resulted in no substantial bias for both of these level-2 variances. Estimates for *Model 2*, however, resulted in substantial positive bias for all conditions. That is, when the level-2 variance component, $\tau_{0(j_1 X j_2)\text{k}}$, was generated to be non-zero, but was *not* estimated in the estimating model (*Model 2*), the remaining level-2 variance components, τ_{0j_10k} and τ_{00j_2k}, associated with the cross-classified factors, were being *overestimated*. In some of the simulated conditions, this overestimation was as high as 44.3%. This corresponds to findings in Luo and Kwok (2009), where misspecification of a variance component at level-2 resulted in positive parameter bias for the remaining level-2 variance component. These findings also correspond to Shi et al.'s (2010) findings, where substantial positive bias occurred for the remaining level-2 variance components, whenever the level-2 variance, $\tau_{0(j_1 X j_2)\text{k}}$, was generated, but not estimated. For this simulation, overestimation of the level-2 variance components was impacted by several of the simulated conditions.

Estimates of the level-2 variances (from *Model 2*) were *more* positively biased for conditions of 2 feeders. This trend was the opposite of what was seen for the level-1 variance estimates. Here, as the data structure became more cross-classified, that is, students were "fed" to more classrooms, overestimations of these two variance components actually decreased. This trend corresponds with Luo and Kwok's misspecification of the three-level CCREM, where overestimation of the remaining level-2 variance components decreased as the structure of the cross-classification increased. Mean bias differences between the two conditions were over 10%, so that when students were sent to 4 classrooms (instead of 2), overestimation decreased by 10.7% for τ_{0j_10k} and 11.5% for τ_{00j_2k}. This impact of the cross-classification structure on bias for the level-2 variance components depended on another condition, the equality of the level-2 variance components.

Estimates of the level-2 variance components (for *Model 2*) were *more* positively biased when the level-2 variance components were all set to be equal in value. Conversely, when the level-2 variances were unequal and, therefore, the variance component $\tau_{0(j_1 X j_2)k}$ was a smaller value, there was *less* overestimation of the remaining level-2 variance components. Mean bias differences between the two conditions were noticeable. When the level-2 variances were all equal (versus unequal) overestimation increased by 24.1% for τ_{0j_10k} and 17% for τ_{00j2k}. Bias differences between equal and unequal decreased when the data structure became more cross-classified. That is, the difference in mean bias between equal and unequal was larger for conditions with 2 feeders than 4 feeders. In other words, as the data structure became more cross-classified, the impact of the equal condition on the estimates for the level-2 variances $\tau_{0j_1 0k}$ and τ_{00j2k} grew smaller.

Shi et al (2010) did not manipulate the equality of the level-2 variance values, but did manipulate the value of $\tau_{0(j_1\ X\ j_2)\text{k}}$, and found that this manipulation had no substantial impact on bias for the remaining level-2 variances associated with the cross-classified factors (when $\tau_{0(j_1\ X\ j_2)\text{k}}$ was assumed to be zero). As observed here, this condition of equal/not equal did impact the level-2 variance estimation bias. It may be the case that the relationship between the value of $\tau_{0(j_1\ X\ j_2)\text{k}}$ and the other level-2 variance components is the underlying factor that is affecting the bias, and not just the value of $\tau_{0(j_1\ X\ j_2)\text{k}}$, independently. Additionally, Shi et al did not vary the cross-classification of the data structure, which was here seen to be an impacting factor on the overestimation of the level-2 variance components.

Although some bias was observed for the SE associated with the level-2 variance components, none of it was substantial. When rounded to the nearest tenth decimal place, estimates were similarly biased for both *Model 1* and *Model 2* within the same condition. That is, the simulated conditions had similar impacts on the SE estimates associated with the level-2 variance components. Shi et al. (2010) had similar results, with very little impact on standard errors when $\tau_{0(j_1\ X\ j_2)\text{k}}$ was generated, but not estimated. Further investigation is needed to determine under what conditions the underestimation of the standard errors may be impacted.

Level-2 Variance (interaction), $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$

Overall, substantial bias was not found for the estimation of the level-2 variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. Bias was only calculated for *Model 1*, because *Model 2* did not estimate this component by design. That is, when the variance component $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was generated *and* estimated, no substantial bias was observed in any of the manipulated conditions. Similarly, no substantial SE bias was found for the standard error associated with $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ in any conditions. Although bias for the standard error existed, neither of the four conditions that were manipulated had any significant effect on the bias that was calculated. This was similar to Shi et al.'s study (2010), in which none of the manipulated conditions impacted the standard errors of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

One of the purposes of the simulation study was to investigate how well estimates of the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, could be recovered under conditions where there may be small within-cell sizes. Often times, this parameter is assumed to be zero because small within-cell sample sizes can make estimation problematic (Meyers & Beretvas, 2006; Raudenbush & Bryk, 2002, Shi et al., 2010) Raudenbush and Bryk state that small within-cell sample sizes may make it difficult to differentiate the variance, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, from the level-1 variance, resulting in estimation problems. Raudenbush and Bryk do not provide a definition for small within-cell sample sizes, other than referencing their dataset, which had within cell sizes as small as 1. Average within cell sizes for the present study ranged from 4 (for small classroom sample sizes with 4 feeders) to 12 (for large classroom sample sizes with 2 feeders). The present simulation did not encounter any estimation problems across any of the conditions. Estimation of $\tau_{0(j_1\ X\ j_2)\text{k}}$ was not problematic when average within-cell sizes were as small as 4 students. Additionally, no substantial bias for the level-1 variance component, $\sigma^2_{i(j_1,j_2)k}$ nor for $\tau_{0(j_1\ X\ j_2)\text{k}}$ was encountered when they were both included in the estimating model (as in *Model 1*). Both of these findings are somewhat surprising, based on the expectations presented by Raudenbush and Bryk (2002).

These findings are only for the scenarios we are investigating. Other scenarios, including the estimation issues discussed by Raudenbush and Bryk may be encountered when the conditions are not similar. Future research needs to investigate different manipulated conditions where small within-celll sizes may present estimation problems. It may be the case that some of the factors in the study (three-level models, cross-classification structure, values of the variances, etc.) may contribute to better parameter recovery, and the effect of small within-cell sizes was therefore not observed here. As more information is found regarding what factors might affect the estimation of $\tau_{0(j_1\ X\ j_2)\text{k}}$, it can become more clear what the best method would be in handling the estimation of this variance with smaller within-cell sizes. The findings here are only for the conditions investigated and may change under different scenarios.

Total Level-2 Variance

A single bias value was calculated using the three level-2 variance components that were included in the simulation study: the first cross-classified factor's variance, $\tau_{0 j_1 0k}$, the second CC factors' variance, $\tau_{00 j_2 k}$, and the variance associated with the interaction of these two crossed factors' random effects, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. Aggregating the bias at this level allowed for additional investigation of what was happening with the level-2 total variance when, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was generated, but not included in the estimating model (*Model 2*).

Overall, substantial negative bias of the total level-2 variance estimation was found for 14 of the conditions in the study. Underestimation of the total variance was as high as 16.8%. The condition that affected the impact of the bias the most was the cross-classification of the data structure. Less negative bias (7.8% less) was found for conditions with 2 feeders than with 4 feeders.

Overall trends for the total level-2 variance components were of most interest. Comparing the bias trends across all three levels allowed for a better depiction of what may be happening with the variance of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ when it is generated, but not estimated. Luo & Kwok (2010) found that when variance related to the cross-classification factors at level-2 was not appropriately modeled, the variance would be "redistributed" to the variance at levels above and below where the cross-classification occurred. Here, the simulation study found that when the variance $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is generated, but not estimated, the variance at level-1 is overestimated, the individual variances associated with the cross-classified factors at level-2 are overestimated, however, the total variance at level-2 is underestimated, and the variance at level-3 is underestimated slightly.

Overall, because the level-2 variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, was generated but assumed to be zero, estimates of the variance at level-1 and for the two variances at level-2 were inflated, while the variance at level-3 was underestimated. The positive bias at level-2, however, was not enough to account for the total level-2 variance that was generated. While some of variance was mostly redistributed to the other level-2 variance components, negative bias for the total level-2 variance still remained, implying that not all of the variance of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was redistributed at just level-2. Some was estimated at level-1, as well. These patterns somewhat follow what was found in Luo and Kwok (2009), where misspecification at the level where the cross-classification occurs resulted in positive bias for the level-1, level-2, and level-3 variance components.

Level-3 Variance, τ_{000k}

No substantial bias was found for the estimation of the level-3 variance for *Model 1* (which estimated $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$) or *Model 2* (which assumed $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ was 0). Overall then, when conditions are similar to those found in this study, and cross-classification is occurring at level-2, using either of the two models to estimate parameters in a given dataset should not result in substantially biased parameter estimates for the level-3 variance component. This contradicts the findings in Luo and Kwok (2009), where overestimation was found for the level-3 variance component, when level-2 was misspecified in the estimating CCREM model. They did not however generate or estimate the variance component, $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$, which may help explain the difference in findings. Shi et al. (2010) did not include a third level in their study, so direct comparisons for these findings are not possible.

Additionally, no substantial relative parameter bias was found for the standard error associated with the level-3 variance component. This finding *is* similar to Luo and Kwok's results (2009), who found no substantial bias for the level-3 standard errors when one of the level-2 cross-classified factors' was not modeled correctly.

Total Variance

Bias calculated for the total variance estimates provided a measure of overestimation or underestimation that may be occurring for the total variance estimates across the three levels in the model. No substantial bias was found for the estimation of the total variance for *Model 2* (which assumed $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ was 0). Although overestimation was occurring for the level-1 and level-2 variance components, it is assumed that $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ is zero and thus the total level-2 variance was under-estimated. Combining that under-estimation with the slight over-estimation of level-1 and level-3 variances resulted in no substantial over bias in the total variance across levels. This implies that despite the model misspecification in *Model 2*, the total variability was recaptured although the unmodeled $\tau_{0(j_1 \ X \ j_2)\mathrm{k}}$ was re-distributed.

Information criteria

Only three information criteria were investigated in the present dissertation, although more are available. Here, values for the AIC, the AICC, and the BIC were compared for *Model 1* and *Model 2*, where the lowest value indicated the best fitting model. Overall, the AIC and AICC consistently selected *Model 1* as the better fitting model. The BIC was not as consistent, and had correct model selection percentages as low as 10.20% in some conditions. Overall, the BIC would select *Model 2* (the misspecified model) in conditions where the IUCC was 7% and the level-2 variance components were not equal. The percentages for correct model selection increased when the classroom sample size was larger.

Only one simulation study has specifically investigated the use of information criteria for model selection of CCREM models (Beretvas & Murphy, 2013). They compared several information criteria in addition to the ones used in this study: Hannon and Quinn's information criteria (HQIC; Hannon & Quinn, 1979) and Bozdogan's consistent AIC (CAIC; Bozdogan, 1987). In their study, they found that the default information criteria for SAS (including AIC and BIC from the present study) were not always the best performing indices for misspecified CCREM models. The high correct model identification of the AIC and AICC in this study warrant further research. Beretvas and Murphy provided corrections that could be made in the calculation of BIC and HQIC to provide more accurate model information criteria. The present study, however did not apply this correction for the BIC and did not investigate the HQIC. Future research should look at the performance of these corrections for similar conditions found in the present study, especially because of the inconsistent model selection that was found for the BIC with smaller IUCC values.

LIMITATIONS AND FUTURE RESEARCH

The present dissertation was an extension of the work conducted by Shi et al. (2010), guided by the findings in Luo and Kwok (2009) concerning intermediate level cross-classification in three-level models. At present, this is only the second study to investigate recovery of the estimates of the variance component, $\tau_{0(j_1\ X\ j_2)k}$, and like most studies, there are several limitations that need to be considered.

Some of the limitations in this study are related to the study's design. For example, in order to generate small within-cell sample sizes, two different conditions were manipulated simultaneously: classroom sample size and cross-classification structure. Specifically, the conditions with smallest average within-cell size (4 students) was the *same* conditions with 4 feeders and small classroom sample size. The successful recovery of the estimates of $\tau_{0(j_1\ X\ j_2)k}$ and its associated standard errors may be confounded by the number of receivers and the classroom sample size. In comparison, the average cell size for the 2 feeder condition and small classroom sample size was 8 students. The successful recovery in small within-cell sample sizes may only occur with 4 feeders, and the problematic estimation of $\tau_{0(j_1\ X\ j_2)k}$ discussed in applied research might start to be observed when the data structure is less cross-classified with an average cell size of 4. Future research should be conducted that explores additional patterns of cross-classification that include a small number of receivers with small within-cell sample size values and investigate the resulting impact on the recovery of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

The present simulation study is also limited by the IUCC values chosen for the random effects' interaction, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. These values were chosen based on common IUCC values found for classrooms within the same elementary school, as well as, those found for the real world data set in the first study. This allowed for a realistic simulation for the cross-classified factors variance component values, $\tau_{0j_1 0k}$ and τ_{00j_2k}. Unfortunately, little research has included the estimation of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, and, although one value was provided from the real world data set, no other values seemed to be available from previous applied research. This makes it difficult to determine what a realistic IUCC value would be for $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. The simulation study conducted above implied that the IUCC value would have an impact on bias, which may create bias issues for applied researchers who assume this component is zero in their dataset and do not estimate it. Future research should investigate other IUCCs values, and explore what may be considered a common value for $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, and what values may affect the ability to produce unbiased estimates for $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

In addition to the limitations listed above, the simulation study was also limited by where the cross-classification was generated to occur. Cross-classification can happen at several levels of a multi-level model, each with its own set of possible estimation difficulties that occurred when the cross-classification is not modeled appropriately. When cross-classification occurred at level-2 or level-3 of a dataset, and it was not modeled correctly, biased estimates were found for parameters in levels where the cross-classification did not occur (like level-1) (Luo & Kwok, 2009). Consequently, the cross-classification may occur at level-3, but level-2 and level-1 estimates may be biased. In the current simulation study, the cross-classification occurred at level-2 and misspecifications related to the cross-classification structure affected estimates at level-1. Future research needs to investigate cross-classification at other levels, such as the third or fourth level, to assess whether estimates of the variances for lower levels in the model are affected.

One final limitation in the present study, concerns the predictors at level-1 and level-2 of the estimating model. For this study, these predictors were modeled as fixed at both levels. This may not necessarily reflect a realistic CCREM model, where a researcher may be interested in allowing these predictors to randomly vary. Introducing this random variation may affect bias in the estimation of the random effects variance components, including the estimation of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. This may especially be the case for predictors at level-2, when the cross-classification is occurring at level-2, as well. Future research should investigate the effects of allowing predictors to randomly vary on estimation of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$.

EDUCATIONAL IMPORTANCE AND CONCLUSION

Educational researchers often encounter data that is inherently hierarchically structured. Appropriate modeling of these hierarchical structures is essential, and these data structure may not always be purely nested. When cross-classification is found to occur in a data structure, the cross-classification of the structure must be appropriately model to avoid biased estimates that may occur when the cross-classification is ignored (see, for example, Beretvas; 2008; Luo & Kwok, 2010; Meyers & Beretvas, 2006. Correctly modeling cross-classification in educational data introduces a variance component that is not often mentioned or estimated in the applied literature: the variance component for the cross-classified factors' random effects' interaction, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. While applied researchers in the education field are now more often modeling cross-classification structures as cross-classified, it is still common practice for applied researchers to assume that $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is zero, and not include its estimation in the model.

Results for the simulated and real world data studies conducted here demonstrated that assuming the value of this variance component is zero, and not estimating it, may lead to overestimation of some estimates across several levels of the model. While the fixed effects and their associated standard errors were not affected with this type of misspecification, if a researcher is making decisions about adding predictors, or calculating the amount of variance explained by a model, more caution may need to be taken when considering a model that estimates $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. It may be the case that the true value of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is not zero, and assuming that this parameter is zero may result in overestimation of the level-1 variance components, as well as overestimation of the level-2 variance components that are associated with the cross-classified factors. If a researcher's primary interest is only interpretation of fixed effects, or standard errors, misspecifying the model by assuming that $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ is zero, does not seem to affect the outcome. If, on the other hand, the researcher is looking to add explanatory variables to their model, or is investigating the percent of variance explained by their grouping or clustering factors, more caution should be practiced when deciding whether to include estimation of $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ in the model.

Overestimation of the level-1 variance components can lead the researcher to draw incorrect conclusions about the data set. Researchers can base decisions about adding student-level predictors to the model using the value of the level-1 variance component value. New predictors might be added to the model to explain some of the level-1 variance component, however, that variance might better be explained by the level-2 variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$. There may be an effect of the cross-classified factors interaction that has not been considered, and might be incorrectly attributed to the student level variability. By modeling the variance component, $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$, researchers can correctly attribute variability associated to the random effects interaction and not attribute it (possibly incorrectly) to the student level variance.

The overestimation of the level-2 variance components associated with the cross-classified factors may also impact researcher decisions. By not modeling $\tau_{0(j_1\ X\ j_2)\mathrm{k}}$ when it may exist, the variance associated with the two cross-classified factors will be inflated. Inflation of these two factors may lead to incorrect conclusions about how much each of these factors contributes to the total variability in the data. In the case of the cross-classified factors being classrooms, overestimation of the classroom effect can mislead a researcher about how the classroom impacts an outcome of interest. Appropriately modeling the variance component $\tau_{0(j_1 \ X \ j_2)k}$, when it exists, would provide less biased estimation of the level-2 variance component.

To offer a simple example that can be encountered, researchers might be interested in investigating outcomes for a specific grade level, like first grade. If the researcher has access to the student's data from the previous year, they may be interested in modeling the hierarchical nature of this data structure. The students in the first grade classrooms may not have all attended the same kindergarten classes, and vice versa, so the researcher is presented with a cross-classified data structure, which would include pure nesting at level-3, elementary school, if the students in these classrooms stayed in the same school.

To appropriately model this example data, the researcher might opt to use CCREM models like *Model 1* or *Model 2* in the simulation study. Based on findings in the applied literature, it would be likely that the researcher would not include the estimation of $\tau_{0(j_1 \ X \ j_2)k}$ (here, the interaction of the random effect associated with the kindergarten classroom and the random effect associated with the first grade classroom). Instead, the researcher would probably apply a model similar to *Model 2*. As demonstrated in the simulated study, however, by not estimating the variance, $\tau_{0(j_1 \ X \ j_2)k}$, when it actually exists, there is an increased chance that the estimates for the two classroom variances and the level-1 variance will be overestimated.

This overestimation can be further affected by conditions where there are different cross-classification structures for the dataset, such as if very few students are considered cross-classified, or conditions where the IUCC of the classrooms are larger. The suggestion here would be for the researcher to include the estimate of this variance component in their model (much like *Model 1*), even if the within-cell sizes are smaller. Including $\tau_{0(j_1 X j_2)k}$ in the estimating model would increase the likelihood of providing less biased estimates of the level-2 variance components associated with the kindergarten classroom and first grade classroom, as well as the level-1 variance component, associated with the student.

Including this interaction in the model would provide the opportunity for the researcher to interpret and discuss interaction effects between kindergarten classrooms and first grade classrooms. To extend the original example, the data may include several students in the same first grade classroom that may have come from different kindergarten classrooms. If variances associated with kindergarten and first grade classrooms are large, the researcher would likely include predictors in their estimating model that may help explain the variability in these classrooms. These predictors can be variables like teaching experience of the classroom teacher, or even the size of the classroom. It might be of interest to also investigate how the interaction of attending a specific kindergarten and first grade classroom might impact the student outcome. Perhaps, having attended one kindergarten classroom might have more of a positive effect on a student's first grade classroom experience than another student's experience who came from a different kindergarten classroom, but attended the same first grade class. In addition to investigating the variance associated with this interaction, the researcher might look to explaining this interaction with some type of explanatory variable. Since the variance, $\tau_{0(j_1 X j_2)k}$, is associated with the interaction of two cross-classified factors, it would make sense to include predictors associated with each of these factors, as well as the interaction of the predictors, to help explain the variability occurring at level-2.

Overall, substantial parameter bias was found for model estimations when the variance component $\tau_{0(j_1 X j_2)k}$ was assumed to be zero, and was, therefore not estimated. The assumption that this variance component is zero, and, consequently, its exclusion from an estimating model, affected bias of the level-1 and level-2 variance component estimates. Several of the conditions investigated (IUCC, cross-classified structure, level-2 variance component values) also seemed to impact parameter recovery. Given that the variance components for level-1 and level-2 can often be used to make decisions about factors like classroom effects, it is recommended that researchers with cross-classified data structures should include $\tau_{0(j_1 X j_2)k}$ in their estimation model.

References

Achilles, C.M., Bain, H.M., Bellott, F., Boyd-Zaharias, J., Finn, J., Folger, J. Johnston, J. & Word, E. (2008). *Tennessee's student teacher achievement ratio (STAR) project*. Retrieved from http://thedata.harvard.edu.

Ainsworth, J.W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces, 81(1), 117-152*.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), Second international symposium on information theory (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Attalli, Y., & Powers, D. (2009). Validity of scores for a developmental writing scale based on automated scoring. *Educational and Psychological Measurement, 69*(6), 978-993.

Beretvas, S. N. (2008). Cross-classified random effects models. In A. A. O'Connell & D. Betsy McCoach (Eds.), *Multilevel Modeling of Educational Data* (pp. 161-197). Charlotte, SC: Information Age Publishing.

Beretvas, S. N. (2010). Cross-classified and multiple membership models. In J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 313–334). New York, NY: Routledge.

Beretvas, S. & Murphy, D. (2013). An evaluation of information criteria use for correct cross-classified random effects model selection. *The Journal of Experimental Education, 81*(4), 429-463.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52,* 345–370.

De Fraine, B., Van Landeghem, G., Van Damme, J., & Onghena, P. (2005). An analysis of well-being in secondary school with multilevel growth curve models and multilevel multivariate models. *Quality and Quantity, 39,* 297–316.

Enders, C.K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138.

Fielding, A. (2002). Teaching groups as foci for evaluating performance in cost effectiveness of GCE Advanced Level provision: Some practical methodological innovations. *School Effectiveness and School Improvement, 13,* 225-246.

George, R., & Thomas, G. (2000). Victimization among middle and high school students: A multilevel analysis. *The High School Journal, 84*, 48–57.

Goldstein, H. (1995). Multilevel Statistical Models (2nd edition). New York: John Wiley.

Goldstein, H. (2003). *Multilevel statistical models* (3rd Ed.). London: Hodder Arnold.

Goldstein, H. (2010). *Multilevel statistical models* (4th Ed.). Chichester, UK: Wiley & Sons, Ltd.

Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician, 60*, 19–26.

Hannon, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B, 41,* 190–195.

Hedges, L.V. & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29(1),* 60-87.

Hoogland, J.J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research, 26(3), 329-367.*

Hox, J. J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Lawrence Erlbaum Associates.

Hox, J. J. (2010). Multilevel analysis: Techniques and applications (2nd Ed.). New York: Routledge.

Hox, J. J. & Kreft, I. G. (1994). Multilevel analysis methods. *Sociological Methods & Research, 22*(3), 283-299.

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika, 76,* 297–307.

Lloyd, J., Li, L., & Hertzmann, C. (2010). Early experience matters: lasting effect of concentrated disadvantage on children's language and cognitive outcome. *Health & Place, 16*(2), 371—380.

Luo,W. & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44,*182-212.

Luo, W. & Kwok, O. (2012). The consequences of ignoring individuals' mobility in multilevel growth models: A Monte Carlo study. *Journal of Educational and Behavioral Statistics, 37*(1), 31-56.

Ma, X., & Ma, L. (2004). Modeling stability of growth between mathematics and science achievement during middle and high school. *Evaluation Review, 38*, 104–122.

Ma, Xin & Wilkins, J.L.M (2002). The development of science achievement in middle and high school-individual differences and school effects. *Evaluation Review, 26(4), 395-417*.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*, 14–28.

Meyers, J. L. & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*(4), 473-497.

Rasbash, J., & Browne, W. J. (2001). Modeling non-hierarchical structures. In Leyland, A. H. and Goldstein, H. (Eds.), *Multilevel modeling of health statistics* (pp. 93-105). Chichester: John Wiley & Sons.

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics, 19*(4), 337-350.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

SAS Institute Inc. 2012. What's New in SAS® 9.3. Cary, NC: SAS Institute Inc.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modeling. *British Journal of Mathematical and Statistical Psychology, 63,* 1-15.

Spybrook, J. & Raudenbush, S.W. (2009). An examination of the precision and technical accuracy of the first wave of group-randomized trials funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis, 31* (3), 298-318.

Szapocznik, J., Lombard, J., Martinez, F., Mason, C. A., Gorman-Smith, D., Plater-Zyberk, E., Brown, S. C., & Spokane, A. (2006). The impact of the built environment on children's school conduct grades: The role of diversity of use in a Hispanic neighborhood. *American Journal of Community Psychology*, *38*(3-4), 299-310.

U.S. Department of Education, National Center for Education. (2011). Digest of Education Statistics, Table 114.

Whittaker, T. A., & Furlow, C. F. (2009). The comparison of model selection criteria when selecting among competing hierarchical linear models. *Journal of Modern Applied Statistical Methods*, *8*, 173–193.